Northumbria University
NEWCASTLE

Automated Early Prediction of Cerebral Palsy: Interpretable Pose-based Assessment for the Identification of Abnormal Infant Movements

Kevin D. McCay

PhD

2022

Automated Early Prediction of Cerebral Palsy: Interpretable Pose-based Assessment for the Identification of Abnormal Infant Movements

Kevin D. McCay

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy

Faculty of Department of Computer and Information Sciences

May 2022

Abstract

Cerebral Palsy (CP) is currently the most common chronic motor disability occurring in infants, affecting an estimated 1 in every 400 babies born in the UK each year. Techniques which can lead to an early diagnosis of CP have therefore been an active area of research, with some very promising results using tools such as the General Movements Assessment (GMA). By using video recordings of infant motor activity, assessors are able to classify an infant's neurodevelopmental status based upon specific characteristics of the observed infant movement. However, these assessments are heavily dependent upon the availability of highly skilled assessors. As such, we explore the feasibility of the automated prediction of CP using machine learning techniques to analyse infant motion.

We examine the viability of several new pose-based features for the analysis and classification of infant body movement from video footage. We extensively evaluate the effectiveness of the extracted features using several proposed classification frameworks, and also reimplement the leading methods from the literature for direct comparison using shared datasets to establish a new state-of-the-art. We introduce the RVI-38 video dataset, which we use to further inform the design, and establish the robustness of our proposed complementary pose-based motion features. Finally, given the importance of explainable AI for clinical applications, we propose a new classification framework which also incorporates a visualisation module to further aid with interpretability. Our proposed pose-based framework segments extracted features to detect movement abnormalities spatiotemporally, allowing us to identify and highlight body-parts exhibiting abnormal movement characteristics, subsequently providing intuitive feedback to clinicians.

We suggest that our novel pose-based methods offer significant benefits over other approaches in both the analysis of infant motion and explainability of the associated data. Our engineered features, which are directly mapped to the assessment criteria in the clinical guidelines, demonstrate state-of-the-art performance across multiple datasets; and our feature extraction methods and associated visualisations significantly improve upon model interpretability.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my principal supervisor Dr. Edmond Ho, for supporting me in every way possible throughout my studies. He has provided me with constant encouragement and led me through the field of human motion analysis and machine learning. I thank him for his incredible insight, his invaluable advice, his endless support, and perhaps most importantly his enduring patience!

I would also like to thank the team at the RVI in Newcastle, in particular Claire Marcroft, Prof. Nick Embleton, and Patricia Dulson. Without their expert knowledge and guidance I would not have been able to even start the project, and so I deeply appreciate the time they took out of their busy days to help me with this task.

I would also like to thank Dr. Hubert P. H. Shum for his continuous support, guidance and helpful advice. Furthermore, I thank Prof. Wai Lok Woo for supporting me in the latter stages of my PhD, and for his valuable comments and suggestions.

Thanks also to my fellow PhD students from the University of Northumbria, for their encouragement during my studies and the many helpful and constructive discussions we have had. In particular, I would like to thank Daniel Organisciak, Dimitris Sakkos, and Shanfeng Hu for their tremendous support and encouragement.

Finally, I would have never been able to get this far without the amazing support of my friends and family, but most importantly my wife Elizabeth, to whom I owe everything. I thank her for being there to pick me back up when things were at their lowest ebb. My work could not have been achieved without her continued love and encouragement.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others. The work was done in collaboration with *Dr Edmond Ho, Professor Wai Lok Woo, Dr Hubert P. H. Shum, Claire Marcroft, Patricia Dulson, and Professor Nicholas Embleton*.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the *Northumbria University Engineering and Environment Ethics Committee / Health Research Authority and Health and Care Research Wales / Research Ethics Committee* on *21/05/2019*.

I declare that the Word Count of this thesis is 31,680 words.

Name: Kevin D. McCay

Signature:

Date: 26 May 2022

Publications:

Portions of the work presented in this thesis have previously been published in the following academic papers:

- **Kevin D. McCay**, Pengpeng Hu, Hubert P. H Shum, Wai Lok Woo, Claire Marcroft, Nicholas D. Embleton , Adrian Munteanu and Edmond S. L. Ho, A Pose-based Feature Fusion and Classification Framework for the Early Prediction of Cerebral Palsy in Infants. In IEEE Transactions on Neural Systems and Rehabilitation Engineering, Under Review, doi: 10.1109/TNSRE.2021.3138185, 2021.

- D. Sakkos, **Kevin D. McCay**, C. Marcroft, N. D. Embleton, S. Chattopadhyay and E. S. L. Ho, Identification of Abnormal Movements in Infants: A Deep Neural Network for Body Part-based Prediction of Cerebral Palsy. in IEEE Access, doi: 10.1109/ACCESS.2021.3093469, 2021.

- **Kevin D. McCay**, Edmond S. L. Ho, Dimitrios Sakkos, Wai Lok Woo, Claire Marcroft, Patricia Dulson, Nicholas D Embleton, Towards Explainable Abnormal Infant Movements Identification: a Body-part Based Prediction and Visualisation Framework. In: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), 2021.

- **Kevin D. McCay**, Edmond S. L. Ho, Hubert P. H. Shum, Gerhard Fehringer, Claire Marcroft and Nicholas D. Embleton, Abnormal Infant Movements Classification with Deep Learning on Pose-Based Features. In: IEEE Access, vol. 8, pp. 51582-51592, 2020.

- **Kevin D. McCay**, Edmond S. L. Ho, Claire Marcroft, and Nicholas D. Embleton, Establishing Pose Based Features using Histograms for the Detection of Abnormal Infant Movements. In: IEEE EMBC, pp. 5469-5472, 2019.

Chapter 1

Introduction

The recognition, analysis and reconstruction of complicated motion, such as human activity, has been a popular topic of research for many years. In an early study, Johansson et al. [1] conducted a series of experiments which assessed the ability of humans to perceive the movement patterns of living organisms in motion. The study found that the motion of the human body could be represented as a number of elements which, if temporally visualised in specific motion combinations, allowed for the subject to successfully establish the actions being carried out. This research subsequently contributed towards the idea that human action recognition could be undertaken by computers, allowing for the automation of various activities traditionally requiring human input. Since then significant research has been carried out in the field of computer-vision, specifically relating to human action recognition [2].

The automated recognition of human activity has wide-ranging applications including visual surveillance, content-based video indexing, intelligent monitoring, human-machine learning and virtual reality [3]. This evolving technology also has the potential to have a profound impact upon human computer interaction, subsequently affecting the way we interact with the world [4]. The idea of automating these processes has subsequently been an area of interest for researchers in many varied fields, due to the inherent ability of these techniques to streamline traditionally intensive manual operations. In this thesis, we propose that technology such as this should be considered within the neonatal healthcare domain, to aid with the early diagnosis of movement disorders, such as Cerebral Palsy (CP).

CP is a condition which primarily affects movement, posture and coordination, and describes a group of lifelong neurological disorders usually caused by a brain injury occurring before, during or shortly after birth. CP is one of the most common chronic motor disabilities that can occur in infants, with an estimated 1 in every 400 babies born in the UK receiving a confirmed diagnosis each year [5]. CP is a complex neurological condition, and the severity of the symptoms can vary quite significantly [6]. In order to provide the best possible outcome, early diagnosis is seen as a key area of interest as it has the potential to allow for early intervention clinical care [7]. Early intervention clinical care is particularly beneficial as it allows for opportunities to fully take advantage of the neuro-plasticity found in the early stages of the developing infant brain [8]. Whilst early interventions can take a variety of forms, common challenges are found in the prediction and subsequent diagnosis of CP [9].

The pursuit of early diagnosis of CP has subsequently been an active research area, with some very promising results using tools such as the General Movements Assessment (GMA), the Lacey Assessment of Preterm Infants (LAPI) and the Hand Assessment for Infants (HAI). In practice, the ability to apply these assessments is typically dependent upon the availability of fully trained clinicians. Not only is the training required for assessment using these tools considerable, it is also susceptible to observer fatigue, contains a degree of personal subjectivity and is reliant upon a suitable behavioural state of the infant [10].

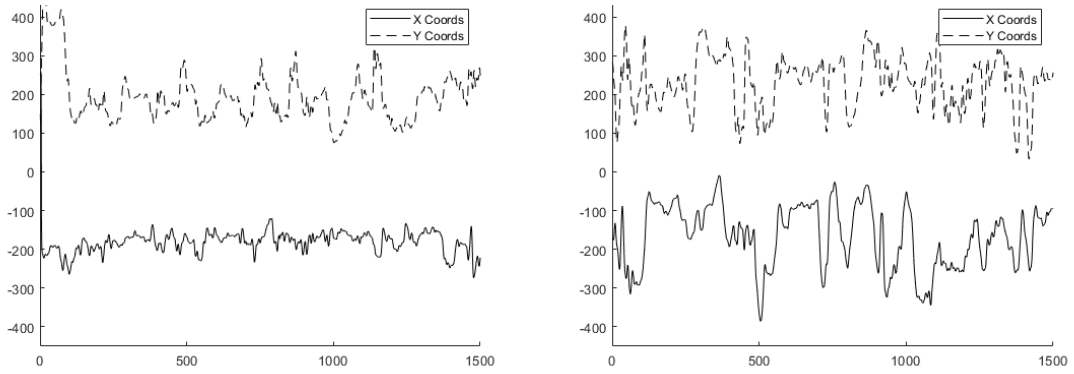
We suggest that there is scope to improve the accuracy, accessibility, and availability of existing diagnostic assessments through computer-based evaluations, to provide quantifiable evidence to clinicians and further information relating to neurological development. We propose that, through the use of state-of-the-art human motion analysis, computer-vision and machine learning techniques, fully automated processes can be developed with suitable accuracy for clinical use. The development of such automated systems will help to significantly reduce the time and subsequent cost associated with current manual diagnostic practices and improve clinical outcomes for infants at risk of CP.

1.1 Key Challenges and Proposed Approaches

In the healthcare domain, there is increasing motivation to utilise technology to aid with clinical decision making, increase predictive accuracy, and target early intervention [11]. In the case of automated CP assessment, there is the potential not only to yield both time and cost savings, but also the opportunity to provide a structure by which clinicians would be able to make earlier and more confident decisions, and a framework which could allow for fully remote diagnostic assessment. Additionally, a machine learning framework could substantiate the decision making process, allowing for intuitive, quantitative, cost-effective, evidence-based assessment [12]. We suggest that by improving automated assessment, such that greater numbers of infants can be assessed using methods based around the GMA, improved outcomes through early intervention care will be evident.

The clinical GMA is seen as the top performing individual examination for the prediction of later CP. This was highlighted in a recent systematic review of the predictive accuracy of assessments to assist in the diagnosis of CP by Bosanquet et al. [13]. In their comparison it was suggested that summary estimates of sensitivity and specificity for the GMA were 98% (95% confidence interval (CI) 74–100%) and 91% (95% CI 83–93%) respectively. This compares well with the other reported methods such as: cranial ultrasound (74% (95% CI 63–83%) and 92% (95% CI 81–96%) respectively), neurological examination (88% (95% CI 55–97%) and 87% (95% CI 57–97%) respectively), and MRI (sensitivity ranging from 86 to 100% and specificity ranging from 89 to 97%). This predictive accuracy, particularly in terms of sensitivity, make the GMA an exceptionally useful tool in the early identification of at-risk infants. This being particularly true, given that the combined sensitivity of assessments such as the GMA and MRI is 100% [14].

By using tools such as automated GMA to help identify the need for further investigation, rather than waiting for children to fail to meet motor developmental milestones, the benefits of early enrichment through neuroplasticity have a greater probability of being seen. In our work, we therefore make use of the GMA, as well as elements from the related Optimality Score, due to the accessibility and simplicity of these tests, as well as their relevance to machine-learning based classification, and the availability of appropriately trained and experienced clinicians to provide ground truth data.



(a) Normal movement (FM+)

(b) Abnormal movement (FM-)

Figure 1.1: Left wrist joint coordinates from two example videos. We observe that the motion representation in the FM+ video shows constant, fluid and controlled movements, rather than the erratic movements shown in the FM- video, aligning well with the GMA.

Several studies have proposed automated solutions to help address the challenges faced, typically making use of frameworks to automatically assess infants based upon the movement patterns associated with the GMA. These methods have been developed over a number of years and now generally incorporate computer-vision techniques to undertake these automated assessments, since these approaches offer greater practicality than physical accelerometers, motion capture systems, and marker-based systems, as discussed in greater detail in Section 2.3. However, due to the limitations inherent in these traditional computer-vision methods, we propose the use of alternate approaches to analyse infant motion, such as automated pose estimation [15, 16, 17, 18] and part-based segmentation [19, 15, 20], since these methods typically represent the state-of-the-art in human activity recognition and motion analysis.

Through the adoption of deep-learning techniques, these new methods of representing human forms from video are typically more robust, accurate, and reliable than previous methods [21]. They are also better able to overcome the previous challenges faced by other methods from the literature, namely, sensitivity to camera movement, object scale, image resolution, self-occlusion, illumination variance, inconsistencies in object size, larger movements between frames, external influences, and manual pre-processing requirements [22]. As such, in our proposed framework, we implement pose estimation as a means of extracting motion data from video footage. Since pose-estimation provides localised joint estimation it is also better able to deal with external influences, meaning that inconsistencies in the footage, like parent/clinician intervention will also be disregarded, as such, there is less requirement for screening the footage prior to analysis.

The localised nature of pose-based analysis can also provide important motion information based upon individual body-part movements (refer to Figure 1.1), requiring comparatively minimal manual tuning [23, 24]. Pose-based analysis also offers lower dimensionality of features, meaning that body movement data can be represented in an abstract manner whilst retaining the most important movement data, contributing to reduced ambiguity in the classification process and greater interpretability.

Given the inherent advantages, our work has therefore focused upon implementing pose-based analysis for infant movement assessment. In our work, we aim to take advantage of these improvements in motion analysis performance by incorporating state-of-the-art methods, such as the OpenPose framework [17], to examine the viability of a pose-based approach for classification. We propose multiple specifically engineered features based upon the GMA, and assess the robustness of each using several machine learning classifiers.

Additionally, since pose-based data is still represented in a format that is human interpretable and simultaneously fully anonymised, this presents an opportunity for data sharing and remote assessment without the requirement for recognisable video data to be transferred, thus ensuring patient confidentiality. Our proposed approach has therefore also allowed us to develop a shareable real-world dataset consisting of labelled skeleton pose data which we have made available to the community.

Finally, given the importance of model explainability within the healthcare domain, we incorporate body-part segmentation as a means of providing additional contextual detail throughout the classification process. Similar to pose-estimation-based methods, body-part segmentation has also been proposed as a means of generating comparable motion information data whilst retaining interpretability. In our case the body is segmented into relevant sections which we suggest can be used for both analysis and visualisation to further improve model interpretability. In our works we examine the feasibility of using the CDCL segmentation method [25] for visualisation such that we can investigate the benefit of this added interpretability for machine learning methods.

1.2 Summary of Contributions

The contributions generated throughout this project can be summarised as follows:

- We propose a series of new pose-based features extracted from 2D video sequences, and engineered based upon specific criteria set out in the GMA, for the analysis and classification of infant body-part movements for the prediction of CP.
- We present the analysis of a new automated feature extraction, fusion, and classification pipeline, for the prediction of CP based upon the GMA. We also make this framework available to the community to further encourage research in this field.
- We propose five deep-learning-based frameworks for the classification of infant body movements based upon our established pose-based features. We make the code and annotated dataset from our deep learning work freely available as an open-source project, to further encourage research in this area.
- We propose a new body-part-based classification and visualisation framework. Our framework uses pose-based features extracted from RGB videos for the spatio-temporal detection of Fidgety Movements (FMs), and is able to highlight pertinent infant body-parts for improved clinical interpretability of machine learning models.
- We present the challenging new RVI-38 video dataset, composed of complex real-world patient data. We make the extracted pose dataset and associated GMA labels available to the community. Given the difficulty in acquiring data in this sensitive area, this is the first real-world clinical dataset made available for this task.
- Finally, we provide experimental re-implementation, comparative evaluation, and discussion of several prominent previous methods, using shared datasets, for unbiased assessment and the generation of a new benchmark. To our knowledge, a comparison of the different proposed methods has not been carried out to quantitatively evaluate the effectiveness of each method on shared datasets.

1.3 Thesis Structure

The remainder of this thesis is structured as follows:

In Chapter 2 we provide an extensive literature review for the task of automated CP prediction. We discuss the background of the project and contextualise the research area by examining the literature relating to the prevalence, diagnosis, and treatment of CP. We then discuss machine learning methods, with a view to how these might be implemented in a CP prediction framework. We then explore the related works and examine several methods which make use of machine learning frameworks for the automated prediction of CP.

In Chapter 3 we provide an overview of our proposed methodology and a discussion of the techniques used, we also discuss how our proposed methods and how are able to overcome several challenges faced by the related works presented in Section 2.3.

In Chapter 4 we provide details of our study design, ethical approvals, data collection procedures, datasets used, selected pose-estimation techniques, and data pre-processessing methods.

In Chapter 5, we provide full details of the feature engineering undertaken to inform our proposed GMA-based features for CP prediction. We also discuss our re-implementation of several prominent methods from the literature to serve as comparative baselines.

In Chapter 6 we provide details of each of our proposed frameworks. We provide information relating to the experimental performance of the proposed features, classification frameworks, and baselines, and how our proposed pipeline evolved throughout the the project development process.

In Chapter 7 we discuss the importance of explainable AI in clinical applications. We examine how we can improve interpretability in the classification pipeline, and propose a new visualisation framework to aid with this task.

In Chapter 8 we provide a conclusion for this thesis and discuss the potential future work in this research area.

Chapter 2

Literature Review

In this chapter we discuss the symptoms of CP, along with the prevalence within different populations, methods of treatment, and the outlook for those affected. We then provide an overview of machine learning methods and examine the leading proposed automated systems for the early identification of CP. Finally we discuss the importance of interpretability in machine learning frameworks within healthcare settings and examine some relevant techniques which have been developed to aid with this, whilst improving upon the performance of previous computer-vision methods.

2.1 Cerebral Palsy

This section introduces the clinical aspects of CP. We discuss the prevalence, outlook, relevant developmental milestones, and the effect that CP has on the life of people diagnosed with the disorder. We discuss the current diagnostic tools for the prediction of CP, the importance of early intervention care, and the different methods of treatment and rehabilitation available.

Cerebral Palsy and the Associated Symptoms

CP is an umbrella term that describes a group of lifelong neurological conditions. CP primarily affects mobility, muscle tone, posture, and coordination, although it can also present a range of other complications, such as difficulties with swallowing, problems with speech-articulation, hearing loss, vision impairment, seizures, gastro-oesophageal reflux disease, and learning disabilities which can contribute towards a reduced ability to learn new skills [26]. The severity of the symptoms of CP can vary significantly, with some individuals presenting with relatively minor symptoms, whilst other are severely disabled and subsequently face great challenges in undertaking routine daily activities [27].

Diagnostically, there are four main categories of CP: Spastic CP, where the muscles are stiff and tight subsequently inhibiting the range of movement and reducing mobility; Dyskinetic CP, where the muscles move involuntarily resulting in muscle spasms and uncontrolled body movements; Ataxic CP, which results in tremors, clumsy movements, and problems with balance and coordination; and Mixed CP which is used to describe someone exhibiting multiple different categories. CP can also be further sub-categorised into hemiplegia, which affects one side of the body; diplegia, in which two limbs are affected; monoplegia, which affects one limb; and quadriplegia, where all four limbs, and often the whole body, are affected [28].

CP is usually attributed to non-progressive damage to the brain in early infancy [26, 29]. Whilst growing in the womb, the developing infant brain can be damaged by incidents such as periventricular leukomalacia; infections caught by the mother during pregnancy, such as cytomegalovirus, rubella, chickenpox, or toxoplasmosis; perinatal stroke; or a direct injury the unborn baby's head. Additionally, CP can be caused by damage to the brain shortly after birth through asphyxiation; infections of the brain, such as meningitis; head injuries; or very low blood sugar levels [30].

CP is one of the most prevalent physical disabilities found in children, with around 1,800 new cases diagnosed every year [27]. A recent study [31], supported this figure by statistically reporting that there are 2.11 diagnoses of CP per 1000 live births, however their research also suggests that there is a significant correlation between an increased prevalence of CP and infants born prematurely. This observation was supported in another study [32], which reported diagnosis rates in the order of 32.4 diagnoses per 1000 infants born very preterm (28-32 weeks gestation), and 70.6 diagnoses per 1000 infants born extremely preterm ($<$28 weeks gestation).

Increased risk of CP is also associated with other factors, such as, babies born with low birthweight, babies being part of a multiple birth, or the mother smoking, drinking alcohol or taking drugs during pregnancy [30]. It is also suggested that whilst the continual development and enhancement of neonatal care has provided a significant decline in infant mortality rates, this has also contributed towards an increase in the incidence and associated severity of cerebral palsy [33]. An indication of this is provided through the figures relating to infant survival rates and the associated rate of severe disability. In the UK, one-in-ten of those surviving at 26 weeks gestation (survival rate 80%) will have a severe disability, with this figure increasing to one-in-four of those surviving at 23 weeks gestation (survival rate 40%), and rising again to one-in-three of those surviving at 22 weeks gestation (survival rate 30%) [34, 35]. As such, multidisciplinary research is ongoing into methods which can provide an early diagnosis and subsequently, lead to enhanced early intervention care.

Early Intervention

In order to provide opportunities for the best possible outcome for an infant's development, early diagnosis of CP is considered essential. This is largely because an early diagnosis can lead to early intervention care which is particularly important for those with emerging and diagnosed CP [36]. Early interventions look to optimise the neuroplasticity of the developing infant brain, thereby inhibiting the impact of impairment. This subsequently increases the infant's short and long term functional ability by minimising the potential development of associative conditions [37].

Additionally, early diagnosis offers an improved clinical structure for treatment [37]. It also provides a more efficient deployment of the associated social, educational and parental support resources, which often rely upon a diagnosis [38]. However, early diagnosis can be difficult and

time consuming, with a confirmed diagnosis rarely made before 18 months of age [9]. This difficulty in providing an early diagnosis is problematic, as the opportunity to capitalise upon the neuroplasticity of the infant brain becomes progressively more limited the later a diagnosis is made [39]. Clinicians and researchers are therefore pursuing methods whereby the conditions causing movement difficulties can be accurately and efficiently diagnosed earlier [10].

Typical signs that an infant may have some form of cerebral palsy include:

- delays in reaching standard developmental milestones;
- hypotonia, where the infant muscle tone can make the infant appear stiff or floppy;
- muscular weakness in the arms or legs;
- jerky or clumsy movements;
- muscle spasms;
- walking on tiptoes;
- tremulous hand movements.

However, many of these signs can be difficult to determine at an early stage in an infant's development, leading to a desire for the development of targeted assessments [30].

Physical examinations of motor control and coordination, such as the General Movements Assessment (GMA) [40], the Lacey Assessment of Preterm Infants (LAPI) [41], the Hammersmith Neurological Examinations [42], the Hand Assessment for Infants (HAI) [43], the Bayley Infant Neurodevelopmental Screener (BINS) [44], and the Segmental Assessment of Trunk Control (SATCo) [45] have subsequently been developed to identify the emerging signs of CP.

Additionally, neurological imaging techniques, such as Magnetic Resonance Imaging (MRI), Computerised Tomography (CT), Electroencephalogram (EEG) and Cranial Ultrasound, are employed to look for lesions or abnormalities in the brain and to monitor brain activity, as a means of providing a confirmed diagnosis of CP [30].

2.1.1 Physical Examinations

The means of providing reliable early diagnosis of CP has been investigated for a number of years, with several physical examinations producing promising results in identifying the emerging signs of CP. Physical examinations typically evaluate muscle tone, interaction, and coordination, as well as the quality, complexity and fluidity of an infant's spontaneous movements at specific windows in their development. Here, we discuss serveral of the most prominent physical examinations currently used as part of both routine clinical care, and neuro-developmental research.

General Movements Assessment

Prechtl's General Movements Assessment (GMA) [46], is a non-invasive and non-intrusive physical examination, developed for the identification of infants exhibiting signs of the neurological anomalies associated with CP. It has a high predictive power for the neuro-developmental outcome of preterm and term infants, subsequently enabling the early detection of infants who are at increased risk of CP, neurological deficits, autism spectrum disorders, or impairments to cognitive function [47].

The assessment is based upon the gestalt visual perception of infant movement patterns, known as General Movements (GMs). GMs are spontaneous, with a complex repertoire, fluidity between transitions, and a controlled quality to the movements. GMs encompass the whole body, manifesting in a diverse range of arm, leg, neck, and trunk movements with specific spatio-temporal motion sequences, which vary in intensity over time [48]. Rotations and frequent minor directional deviations of motion make GMs appear intricate, smooth and elegant. GMs can be recognised in a developing infant from early fetal life through to approximately 4 to 5 months post term, at which point infants start to exhibit intentional, anti-gravity movements [49].

Whilst GMs are present throughout this early period of development, they can be further subcategorised based upon the developmental age of the infant. From term age until approximately 8 weeks post-term GMs assume the form of 'Writhing Movements' (WMs), in which abnormal movements are further classified as as poor-repertoire, cramped-synchronized or chaotic. As the infant ages, WMs begin to dissipate and are replaced by movements which are more fidgety in nature. These 'Fidgety Movements' (FMs) are observable from around 8 weeks to 20 weeks post-term and typically form the basis of the GMA [39].

In a typically developing infant, FMs wax and wane in intensity, speed, frequency, amplitude, and range of motion. There should be notable fluctuations to the rotation, orientation, and displacement around the limb axes [50]. FMs should also have sufficient duration to be observed properly, and have a consistency to their appearance [40]. Conversely, abnormal GMs are identified by the absence of FMs, with a lack of duration, variability, and complexity throughout the movement sequence [51]. Based upon these observations, Prechtl's 'Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants' [52] formed the foundation of the GMA by exploring the specific composition of infant GMs in detail. Prechtl suggests that, in the case of infants with an impaired nervous system, GMs lose their complex and variable character, becoming less fluid and smooth. FMs are seen as a reliable indicator of brain function, and as such, the presence of abnormal GM patterns is seen as a strong predictor that the infant will receive a confirmed diagnosis of CP in later life [52], consequently allowing for abnormal FM patterns to be identified and subsequently classified [48].

The GMA is carried out by analysing video recordings, ideally captured between 12 and 14 weeks post-term. The video is filmed from above, with the baby lying on their back in a supine position. In the videos, the baby should be lightly dressed, not interacting with any objects such as a dummy or toys, and in a calm state. The baby's spontaneous movements are filmed for 3 to 5 minutes, ensuring that their hands and feet are fully in shot for the full duration. The videos are then examined by trained assessors, who aim to identify the abnormal motion patterns associated with neurological impairment [53].

The GMA has evolved over a number of years and has recently been identified as a leading method of predicting later CP, reporting a sensitivity of 98% and a specificity of 94% [54]. It has also proven highly reliable and transferable, with inter-scorer reliability values of 89-93% [55]. Indeed, it has been reported that the GMA outperforms other methods, such as cranial ultrasound and neurological examination, by producing more consistent and reliable individual results [13]. This places the GMA at the forefront of early detection of neuro-developmental impairment, providing evidence to allow for early intervention care and a subsequent reduction in associated sequelae[47].

Lacey Assessment of Preterm Infants

The Lacey Assessment of Preterm Infants (LAPI) [56] is a tool which was developed for clinical use as part of a longitudinal assessment to monitor an infant's development over an extended period of time, and to identify features which may indicate an increased risk of neuro-developmental abnormalities. Changes to clinical practice have lead to earlier discharges, and as such prolonged, repeated assessment becomes impractical, as such the LAPI has been evaluated for its diagnostic capacity at predicting normal motor outcome or CP prior to term age [41].

Similar to the GMA, an evaluation of the infant's motion characteristics are carried out. Movements and responses are grouped in supine, prone, supported sitting, and stance reaction. The assessment requires the completion of specific documentation which evaluates muscle tone (symmetry and caudocephalic maturation of limb) and spontaneous activity (varied movements vs repeated synergistic patterns). This documentation is then used to generate an overall classification of 'normal, 'unusual' or abnormal' based upon a scoring system set out in the documentation checklist.

Research suggests that the LAPI is capable of providing good predictive accuracy (86% sensitivity, 83% specificity, and 96% negative predictive value for subsequent cerebral palsy) for infants assessed $>$ 33 weeks post-mentrual age [41], but is less accurate when applied to infants $<$ 33 weeks post-menstrual age. However, although it is widely used in clinical practice and validation data is limited to the original study [57], the LAPI is currently the only clinical assessment tool designed to assess preterm infants prior to term corrected age [6].

Hammersmith Neurological Examinations

The Hammersmith Neurological Examinations consist of 2 physical examinations used to identify signs of neuro-developmental abnormalities; The Hammersmith Neonatal Neurological Examination (HNNE) and The Hammersmith Infant Neurological Examination (HINE). Due to their accessibility and inter-observer reliability, both the HNNE and the HINE are used throughout the world in clinic and for research.

The HNNE, is an exam which can be used up to the age of 3 months post-term to record neurological findings and to define normality and outcome predictions, but is not currently used for the de-

Figure 2.1: An example of the the Hammersmith Neonatal Neurological Examination [42]

tection and prediction of later CP. It encompasses 34 items assessing tone, motor patterns, observation of spontaneous movements, reflexes, visual and auditory attention and behaviour [42].

The HINE is based upon the same principles as the neonatal exam, and consists of 26 items that assess different aspects of neurological function: cranial nerve function, movements, reflexes and protective reactions and behaviour, as well as some age-dependent items that reflect the development of gross and fine motor function [58]. The HINE is typically used in infants between the ages of 3 and 24 months post-term to identify children at risk of CP, but can also provide information on the associated type and severity of the motor sequelae. Additionally, the exam incorporates supplementary aspects related to neurological function, such as cerebral visual impairment or feeding abnormalities, allowing for an improved definition of the severity of CP, not limited to motor impairment.

The HINE has also seen the development of an 'optimality score' document [58], which is based upon the frequency distribution of scores found in a normal, low-risk, typically developing population of infants. By scoring each of the items separately from 0 to 3 and adding each score together a global metric is calculated that represents the degree of optimal movements demonstrated. In doing so a series of thresholds have been established that indicate the infant's risk of developing difficulties with neurological function and the associated severity. The scoring criteria suggest

that the lower the score the more severe the impairment. The optimality score sheet is useful for identifying specific movement patterns and characteristics associated with later CP, however, it is stated that the HINE which it is based upon is generally only able to predict later CP using scores from assessments after 5-6 months of age [59].

Hand Assessment for Infants

In the case of infants with unilateral CP, there is evidence to suggest that a different focus is required than for those with bilateral CP [37, 55]. In these infants, asymmetric hand-use is commonly one of the first distinguishable signs of CP, as such the Hand Assessment for Infants (HAI) was developed to detect and quantify the hand function of developing infants from 3-12 months of age. The HAI was designed as a means of (1) identifying early clinical signs of CP, (2) predicting outcome, (3) following development, and (4) evaluating early intervention approaches [43]. The HAI measures the degree and quality of goal directed actions of both hands and quantifies their usage, in doing so clinicians are able to evaluate bimanual hand use, or generate a comparison of each hand separately. The test produces a separate score for each hand, subsequently allowing for a quantification of possible asymmetry which is expressed as an asymmetry index representing general upper limb ability.

A preliminary version of the HAI was created which uses 31 test items to generate an overall asymmetry score. These 31 test items consist of 13x2 unimanual test items, where each hand is given a separate score, and 5 bimanual test items, where the combined use of both hands is scored. Each item is rated between 0 and 2 inclusively, with higher numbers representing more advanced ability. As such, quantifiable differences between the hands are established allowing for the asymmetry index to be reported as percentage, with a greater value indicating a larger asymmetry. Finally, the sum of all items is also used to produce a ‘both hands sum score’. The HAI has been demonstrated to be a promising tool which can detect and measure asymmetries, and allow for quantification of development over time for the evaluation of early intervention on hand function [60].

Bayley Infant Neurodevelopmental Screener

The Bayley Infant Neurodevelopmental Screener (BINS) is a screening tool developed explicitly for use in the high-risk infant population. The BINS is carried out by trained assessors who observe specific tasks based on the Bayley Scales of Infant Development [61]. The BINS tool addresses several components associated with developmental ability to evaluate neuro-developmental risk, namely: neurological functions/intactness (posture, muscle tone, movement, asymmetries, abnormal indicators); expressive functions (gross motor, fine motor, oral motor/verbal); receptive functions (visual, auditory, verbal); and cognitive processes (object permanence, goal-directedness, problem solving) [24].

The BINS consists of 11-13 assessment items for different age levels, with each item being scored as 'optimal performance' or 'non-optimal performance' using a pre-defined set of rules. The sum of the non-optimal items in a category provides a classification, with associated thresholds of low-risk, moderate-risk, or high-risk relating to the likelihood of developmental delay [44]. Additionally, the inclusion of tone and quality of movement items, such as ratings of active and passive tone in the upper and lower extremities, and scoring of quality of movement of the upper and lower extremities, differentiates this screening test from other developmental screening tests [62].

In a validation study, [63] reported that the BINS shows a moderate predictive validity (67%–76%) at identifying lower functioning infants, and that it is a satisfactory screening tool for low birth weight infants, but only when used alongside other known biological and social risk factors.

Segmental Assessment of Trunk Control

Throughout the typical development of an infant, the ability to control sitting balance emerges between the ages of 2 to 9 months. This development is characterised firstly by the ability to control head movements, followed by progressive improvements in trunk control. However, in children with neuro-developmental conditions sitting balance is delayed and, in cases where disability is more severe, sitting balance may be affected throughout their lives. As such, it is suggested that an accurate evaluation of a developing child's head and trunk control is essential in understanding the controlled sitting ability for children with CP [45].

In order to comprehensively assess the seated postural control of children the Segmental Assessment of Trunk Control (SATCo) was developed. Based upon the biomechanics of control of vertical trunk posture, the assessment of trunk control is considered by evaluating the many sub-units that must be coordinated to achieve control when sitting. The assessment includes tests of static, active and reactive control, through progressively changing the 'level' of trunk support, from a high level of support at a shoulder girdle to assess cervical (head) control, through support at the axillae (upper thoracic control), inferior scapula (mid thoracic control), lower ribs (lower thoracic control), below ribs (upper lumbar control), pelvis (lower lumbar control), and finally, no support, in order to measure full trunk control. For each trunk segment level static, active and reactive control are scored as present, absent or not tested. Static control is scored as present if the child is able to maintain a neutral trunk posture above the level of hand support; active control is scored as present if the child can maintain a neutral posture during head movement; reactive control is present if the trunk above the support remains stable during external influence i.e. a small nudge. As the degree of support is reduced by lowering the segment levels the number of unsupported free joints increases, subsequently increasing the amount of voluntary control required.

This systematic assessment enables an in-depth analysis of a child's trunk control abilities and, subsequently, allows for early intervention treatments of deficiencies of trunk control [64]. The validity of the SATCo has been evaluated, with reports suggesting reliability of $>.80$, and whilst it has been shown to be a reliable and valid clinical tool, it is geared more towards older children and those undertaking rehabilitative care rather than infants, and as such is not generally suitable for the earliest diagnosis [65].

2.1.2 Neurological Imaging Tests

In addition to physical examinations, neurological imaging tests may be used in the case of infants considered at high risk of developing CP. Imaging tests are typically used to help assess conditions that occur alongside CP, such as seizures, in order to determine their cause. In practice, imaging tests are often used in conjunction with physical tests, to conclusively diagnose CP and to rule out other conditions . The imaging technologies commonly used as part of clinical care to detect and interpret the severity of brain injury consist of Magnetic Resonance Imaging (MRI), Computerised Tomography (CT), Cranial Ultrasound, and Electroencephalogram (EEG) [30].

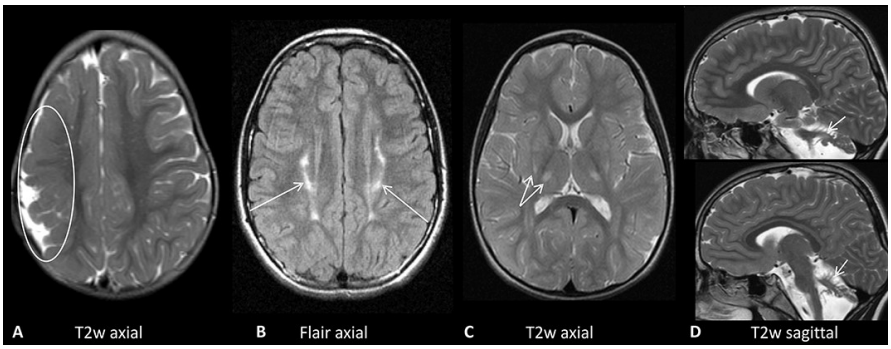


Figure 2.2: Example MRI patterns: (A & B) spastic CP, (C) dyskinetic CP, (D) ataxic CP [66]

Magnetic Resonance Imaging

MRI is a type of brain scan that makes use of magnetic fields and radio waves to produce a detailed three-dimensional image of the brain. In the case of CP diagnosis, these scans are typically used to check for neurological abnormalities in children who may already be be exhibiting symptoms correlated with CP, and can help to establish the associated causes [67]. Scans usually take approximately 1 hour to complete, and require the infant to be sedated throughout the procedure, as it is necessary for them to remain as still as possible throughout the scan. The MRI scanner is a large and expensive piece of equipment, requiring specialist radiographers to carry out the imaging investigations [30], as such it is generally only used on those at high risk of CP, or those awaiting a confirmed diagnosis, and is therefore not routinely used as a screening tool.

Computerised Tomography

A CT scan creates detailed images of inside the body by using a series of x-ray pictures. In this context the CT scan takes cross-sectional images of the brain to provide a means of detecting and diagnosing CP or eliminating other conditions which may present similar symptoms. The imagery generated allows clinicians to identify abnormalities in the brain such as bleeding, skull fractures, lesions or other anomalies that may have led to the development of CP. The scan is painless and typically takes around 20 minutes, however as with the MRI scan the infant must remain still throughout, and as such a mild sedative is administered. The scan is again carried out by trained radiographers using specialist equipment, however in this case ionizing radiation is used when performing the tests, which studies suggest may have a negative impact on developing brains [68]. As such, CT scans are also not used for screening as the risks associated with this procedure outweigh the potential benefits.

Cranial Ultrasound

Cranial ultrasounds are carried out using a small handheld device which emits high frequency sound waves, which can be used to generate images of an infant's brain [30]. Cranial ultrasounds are able to show bleeding in the brain, or subtle changes in white matter, which can subsequently lead to a diagnosis of CP. Whilst the images generated aren't as detailed as those produce by MRI or CT, the system is widely used due to its portability and ease of use, and unlike other methods there is no negative impact upon patients, such as exposure to radiation [69].

Electroencephalogram

An EEG can be used to help diagnose and monitor a number of conditions affecting the brain, and may help to identify the cause of some symptoms commonly associated with CP, such as epilepsy. During an EEG, small pads which can detect the brain's electrical activity are placed upon the scalp. This electrical activity is then assessed to look for abnormal changes in the brain waves which are synonymous with CP symptoms [70]. However, this test is typically used on those at high risk or already suspected of having a brain injury, it is not useful as a screening tool and is typically used to detect epilepsy which, whilst common, is still only present in 30–40% of all cases of CP [71].

2.1.3 Treatment and Rehabilitation

Once a confirmed diagnosis has been made, typically through a combination of physical and neurological examinations [7], a treatment plan can be established by the associated healthcare team. Given that there is no cure for CP, the treatment plan primarily focuses upon helping to ensure that the patient is able to grow to be as active, independent, and functional as possible; subsequently improving the quality of life of both the child and their family. By receiving an early diagnosis, early intervention care can be implemented which looks to optimise motor, cognition, and communication skills, through a multi-disciplinary approach [37, 72]. These interventions typically attempt to manage weakness, spasticity, cognitive dysfunction, nutritional issues, and seizures through the application of specific rehabilitation strategies and the application of evidence-based methods. Research suggests that brain plasticity is integral to the successful treatment of CP, and as such, studies examining rehabilitation therapies and functional recovery techniques have been

developed with the aim of capitalising upon this neuroplasticity [8].

For children and infants, treatment primarily aims to increase mobility and promote physical development, so that milestones can be met, such as sitting, crawling and walking. The associated treatments are started as early as possible and usually continue on a regular basis [73].

For motor and cognitive skills, rehabilitation through physical occupational therapy, and physiotherapy interventions are typically carried out. These interventions use child-initiated movement, task-specific practice, and environmental adaptations to stimulate independent task performance. Additionally, interventions are also carried out to try to prevent secondary impairments and to minimise associated complications [37]. These physical rehabilitative techniques are supplemented by pharmacological therapy to help with ongoing pain and procedural pain, as well as muscle stiffness, swallowing, sleeping difficulties, constipation, drooling, and epilepsy [74].

In some cases surgery is also required to aid with movement difficulties or other CP related issues [30]. Additionally, technological advances such as virtual reality, integrated circuits, wireless communications, physiological sensing, and electromyogram-initiated devices, have been proposed to promote neurological and physical rehabilitation in patients with CP [8].

Studies suggest that these treatment plans should to be established as early as possible and continually reassessed based upon the needs of the individual, as they may require different care and support as they grow older. Specialist knowledge of the cortico-spinal system and the associated neuroplasticity development throughout life is essential for the creation of stage-specific rehabilitation strategies. Combined with early diagnosis, evidence-based therapeutics are key to optimising the early stages of the associated treatment plans, and as such can significantly lessen the severity of afflictions associated with CP [8].

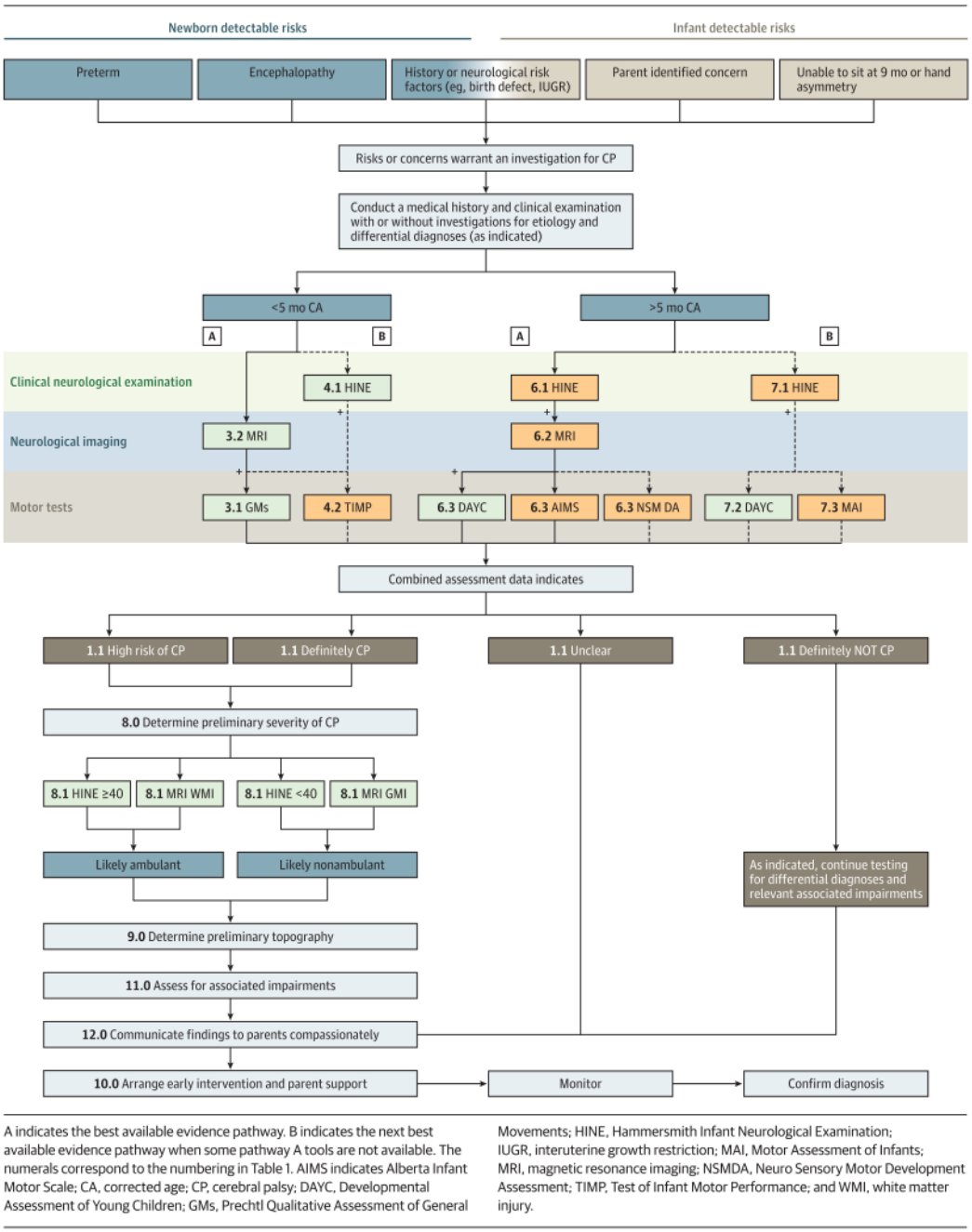


Figure 2.3: Example algorithm for the early diagnosis of CP and an initial treatment plan [37]

2.1.4 Summary

CP is a complex condition with many associative sequelae and as such, management, treatment and rehabilitation are of the utmost importance. We have discussed the value of early diagnosis as a means of providing early intervention clinical care, and the benefits associated with this.

We have examined several prominent methods for diagnosis using both physical examination and

neurological imaging. It is clear that these diagnostic tools have the capacity for accurate diagnosis, typically through a combination of examinations [7]. However, the exams are generally only carried out on those considered high risk, largely due to the associated logistical cost of using these tools as a method of screening for abnormalities. It is conceivable that, given the accuracy, simplicity and accessibility of many of the physical examinations, a screening tool could be utilised to identify all those at risk of developing CP at an earlier stage, allowing for further testing to be undertaken and a confirmed diagnosis reached sooner.

However, whilst physical examinations like the GMA have the potential to provide opportunities for use as a screening tool, the application of these can be challenging, chiefly due to the availability of appropriately skilled clinicians. These clinicians require significant training, as well as years of practical assessment experience to achieve a suitable level of accuracy [53]. Also, given that these tests are based around gestalt visual perception of movement [40], quantifiable predictions are subsequently difficult to specify and discernible diagnostic features can be somewhat speculative. These clinical tests are also heavily reliant upon the infant being in a suitable behavioural state, making them potentially very time-consuming, which can subsequently lead to observer fatigue [75]. It is for these reasons that researchers have started exploring the potential for the automation of these physical examinations, as a means of providing early identification of those at risk, primarily through the use of machine learning techniques.

2.2 Machine Learning

This section introduces the concept of machine learning by briefly discussing its development, definitions, and applications. We determine what constitutes machine learning and provide a high level overview, by comparing the main learning styles and discuss traditional machine learning methods along with deep learning approaches. We then provide an overview of several common classification algorithms.

2.2.1 Background

Machine learning is an evolving branch of Artificial Intelligence (AI) and computer science, based upon computational algorithms that are designed to replicate human intelligence. By gradually learning from examples and experience in the form of data, the algorithms are subsequently able

improve performance in a variety of tasks. Rather than solving problems step-by-step using hard-coded rules, machine learning interprets data to detect patterns and predict desired outputs [76]. Machine learning has seen use in several varied fields, ranging from pattern recognition tasks, to computer vision, engineering, finance, biology, and medical diagnostics [77].

In some areas, machine learning is now able to exceed that of human performance [78]. The ability of a machine learning system to process, learn from, and carry out complex tasks based upon semantic interpretation of specific data, has meant that there has been a significant increase in its use in areas such as medical image recognition [79]. Given that machine learning systems are capable of improving accuracy based upon experience, these systems are typically used to learn from patterns, predict future activity, or make decisions through the integration of components from computer science, statistics, and data science [76].

Machine learning is generally categorised based upon how the algorithm learns to improve its predictive performance [80]. With this in mind, we examine the use of supervised vs unsupervised learning, before discussing traditional machine learning and deep learning techniques for classification.

Supervised Learning

Supervised learning is a subcategory of machine learning which is defined by the use of annotated or labelled datasets. These datasets are used to train a machine learning algorithm to classify data, or predict outcomes with increasing accuracy, by allowing the model to measure its accuracy against ground-truth labels. The algorithm determines any discrepancy between the labels and prediction through the loss function, which is then adjusted until the error has been minimised. As input data is fed into the system, weights are adjusted accordingly until the model has been optimally fitted to the data. This means that the algorithm can then predict the output values for new unseen data based upon the relationships learned from the previous data [81]. Whilst the majority of practical machine learning cases make use of supervised learning, these models require accurately labelled data, which in some cases can be difficult and time consuming to acquire. As such, systems which are able to circumvent this restriction have been proposed [82].

Unsupervised Learning

Unsupervised learning can be considered the antithesis of supervised learning. In unsupervised learning unlabelled data is fed directly into the model. The model then looks for patterns and underlying structures within the data to establish relationships. The algorithms used here analyse and cluster unlabelled data without the need for human intervention [82]. This process can be used for clustering, dimensionality reduction and association, to group data based upon similarities and differences, to reduce the number of data inputs to a manageable size whilst preserving data integrity, and to form relationships between variables. As such, these algorithms are particularly useful in cases where human experts are unsure of exactly what to look for in the data [83].

However, whilst unsupervised learning is helpful in the correct use-case, it also presents several challenges. Unsupervised learning typically has increased computational complexity due to the high volume of training data required [84]. Additionally, there is an increased risk of inaccurate results unless human intervention is included to validate the results. Finally, there is a lack of transparency as to how the data is grouped, which is particularly troublesome in the medical domain [85]. As such, the studies discussed in this thesis make use of supervised learning methods for classification.

Traditional Machine Learning

In order to carry out tasks such as prediction or classification, machine learning algorithms assess input data, which can be labelled or unlabelled, to detect patterns. This input data typically consists of features extracted from the raw data, in the form of structured matrices. An error function is then used to evaluate the model's prediction. In the case of labelled training data, a comparison is made with this 'ground-truth' data to assess the prediction accuracy. The model is then optimised to determine if alterations can be made to better represent the patterns in the data [86].

In traditional machine learning methods, many different algorithms have been proposed, with each offering differing performance depending upon the type of data and the associated task (for more information relating to specific classification algorithms refer to Section 2.2.2). In order to function correctly, traditional machine learning methods rely upon specific characteristics represented in the raw data to be extracted through feature engineering. The features extracted from the raw input data are typically engineered to represent the desired prediction and fed to the model for

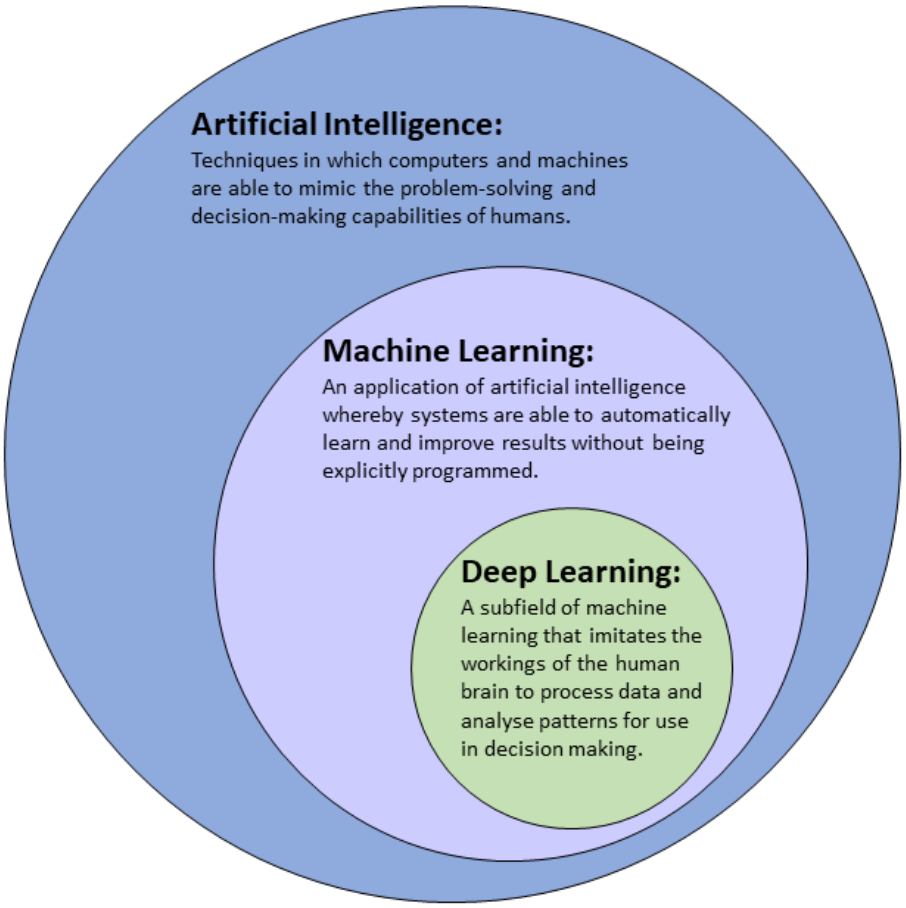


Figure 2.4: The relationship between AI, Machine Learning, and Deep Learning

classification. This process ensures that the input data is in a format suitable for classification using machine learning algorithms, and means that improvements can be made to the model's performance by designing optimal feature representations. This also means that the features are human interpretable, since they have typically been designed to fulfil the needs of the framework output, and as such should represent the relevant diagnostic characteristics.

Deep Learning

Whilst machine learning can be considered a sub-field of AI, so too can we consider deep learning to be a sub-field of machine learning (Fig 2.4). In the case of deep learning, the method of extracting meaningful information from input data differs. Where traditional machine learning largely relies upon engineered features as input, deep learning automates this process, removing the need for human intervention. By eliminating the requirement for feature engineering, deep learning is able to automatically determine the most optimal deep features for the associated task.

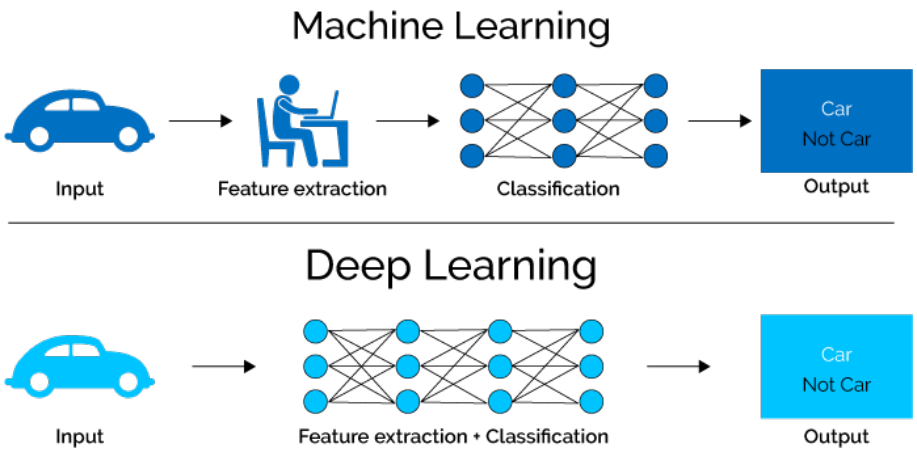


Figure 2.5: Machine Learning vs Deep Learning [90]

Deep learning is customarily associated with the development of artificial neural networks, which were designed to mimic the inter-connectivity of the human brain through layers of nodes. Each node consists of 4 components; an input, a weight, a bias, and an output. The framework calculates the output of each node, and if the output value exceeds a given threshold it activates, passing data to the next layer in the network. An artificial neural network becomes a deep neural network when three or more layers of nodes, inclusive of input and output layers, are incorporated into the framework [86]. This standard architecture feeds the data from the input to the output in one direction, and is called a feed-forward network, however in order to further improve the performance of the framework, backpropogation was introduced and is now commonplace. Backpropogation, passes the data through the network in reverse, moving from output to input, in order to calculate the error associated with each node. This means that the framework can be further optimised, to adjust and fit to the data appropriately [87].

Deep learning can make use of both supervised and unsupervised learning, but generally requires significantly more data to function, since it is does not typically make use of human engineered features. However with more data, we also generally see improved performance in a multitude of machine learning tasks, and as such deep learning represents the state-of-the-art in computer-vision, natural language processing, and speech recognition [88]. However, the use of deep features in deep learning frameworks can cause problems with interpretability, particularly in the medical domain, since the automated decision making process must be transparent and understandable to clinicians [89]. As such, this is an active area of research and is a key consideration when deciding upon a suitable framework for a diagnostic pipeline in healthcare related applications.

2.2.2 Machine Learning Algorithms for Classification

In this section we discuss several of the most commonly used machine learning algorithms and their relevance to binary classification with a view to how they might be implemented within a CP prediction framework.

Logistic Regression (LR)

LR attempts to model probability by determining the correlation between the input independent feature vectors and the categorical predefined class labels by fitting the data to a logistic curve. It is considered to be an extension of linear regression and can be used to solve classification problems and provide the associated probabilities [91].

Support Vector Machine (SVM)

SVM uses supervised learning to analyse data for regression, classification, and outlier detection problems. A hyperplane is generated by the algorithm to separate training data with the maximally permissible margin in high dimensional space. This optimal hyperplane is then used for classification on unseen data by modelling similar patterns contained within the data. Additionally, non-linear classification can be performed using a nonlinear kernel function to replace every dot product, allowing the algorithm to fit the hyperplane within a transformed feature space, subsequently classifying vectors in multidimensional space [92].

Decision Tree (DT)

DT is a non-parametric supervised learning method that determines the class of a given feature vector using a branching node-based structure. Each node of the DT is referred to as either a decision node or a leaf node. The decision nodes consist of two or more branches, and the leaf node represents a classification decision. This method learns conditional control statements which consist of simple decision rules inferred from the data features. The final output is a leaf node which represents the classification of feature vector [93].

Linear Discriminant Analysis (LDA)

LDA is similar to Principal Component Analysis (PCA) in that both are linear transformation techniques typically used for dimensionality reduction. However, where PCA's goal is to find the components that maximize the variance in a dataset in an unsupervised capacity, LDA's goal is to compute the axes that will maximally separate two or more classes using supervision, making it suitable for classification. LDA works using three distinct steps, the first step is to calculate the separability between classes (between-class variance) using the distance between the associated mean of each class, the second step is calculating the distance between the mean of each class and the samples of each class (within-class variance), the third step is to lower the dimensional space to maximize between-class variance and minimize with-class variance [94].

Ensemble of Classification Models (ENS)

ENS methods are learning algorithms which use a combination of multiple other learning algorithms as a means of improving classification performance. The ENS method uses a weighted vote of all the predictions in order to calculate a final prediction. This was originally carried out through Bayesian averaging, but more recently bagging and boosting methods, such as AdaBoost (AB) and LogitBoost (LB), have been implemented to further improve performance. AdaBoost (AB) boosts performance by combining several weak classifiers. It initialises with unweighted feature vectors and raises the weight of the training data for misclassified samples. As such, the next classifier in the sequence is constructed using different weights and misclassified training data gets their weights boosted, with this process subsequently repeated. A majority vote across all classifier predictions are then combined for the final prediction [95]. LogitBoost (LB) further extends AB by using binomial log-likelihood to modify the loss function linearly rather than modifying the loss function exponentially, which helps deal with noise and outliers in the data [96].

k-Nearest Neighbour (kNN)

kNN is a non-parametric classification method which uses similarity measures within the training dataset to classify test data. Essentially, kNN uses the k closest samples that are involved in the majority voting process to predict a data point. kNN typically performs well with data that is small in size, labelled and noise free [97].

Convolutional Neural Network (CNN)

A CNN is a feed-forward deep learning neural network which is most commonly used to analyse visual data. CNNs are widely used in computer vision tasks and have come to represent the state of the art for many visual applications, such as image classification. The construction of a CNN is analogous to the connectivity pattern of neurons in the human brain, with input imagery processed in a manner similar to that of the visual cortex. The CNN uses stacked convolutional layers to convolve input maps, to generate deep feature map representations for classification [98].

2.2.3 Summary

In this section we have discussed machine learning by outlining supervised learning, unsupervised learning, traditional machine learning and deep learning. We have discussed several machine learning algorithms and briefly introduced the concept of interpretable AI, a notion which we discuss in greater detail in Section 3.1.4.

The development of machine learning has provided opportunities for the application of automated decision making processes in several varied fields, including medical diagnostics. In the next section we examine several studies which propose the use of machine learning, specifically for the purpose of predicting later CP based upon the movement data provided through the physical examinations discussed in Section 2.1.1.

2.3 The Application of Machine Learning for CP Prediction

Whilst many of the physical examinations associated with the prediction of later CP have proven to be accurate, non-invasive and non-intrusive diagnostic tools, each requires a significant investment in terms of both time and resources. Additionally, the assessments themselves are subjective, and whilst the diagnostic accuracy amongst experienced assessors is high, the gestalt nature of the assessments make it difficult to quantify discernible diagnostic features manually [40]. The assessments also require the infants to be examined whilst in a suitable behavioural state [75], as such these assessments can prove to be time-consuming, which can therefore lead to observer fatigue [99]. In practice, the challenges of applying these assessments primarily relates to the availability of appropriately trained and skilled clinicians. These clinicians require considerable

additional training and as such, physical examinations are currently only carried out on infants at high risk of developing CP, and to support a diagnosis where there are existing medical concerns; they are not currently used as a screening tool for healthy babies [100].

A system which can quantify an early diagnosis has the potential not only to support the decisions and diagnoses made by healthcare professionals, but the development of a suitably reliable automated tool would mean that the analysis of all babies could be carried out. It is feasible that through the adoption of digitally enabled care, the development of an AI-based automated system could help to alleviate the aforementioned issues; easing current workforce pressures, enhancing care quality, improving patient outcomes and providing cost effectiveness.

2.3.1 Automated Assessment

Automated assessments could conceivably provide measurable evidence to clinicians to support decisions made about care and to track the progress of physical rehabilitation. There is also scope for these systems to help reduce the time and cost associated with current diagnostic practices, and potentially become part of wide scale screening programs.

The importance of reassuring parents that their child is developing normally has also been highlighted, as such quantifiable diagnosis presents an opportunity to achieve this in an interpretable manner. As such, several studies have been carried out which attempt to assess the viability of automating the diagnostic processes used to predict motor impairment, based upon observed motion quality (refer to Table 2.1). In a recent study, Marcroft et al. [10] discussed the use of several technologies which analyse recorded movement data for this purpose. They suggest that the adoption of these technologies could potentially help with the prediction of motor impairment in infants. The study also highlighted the need for the early identification of those most at risk, and proposed that the application of these technologies within neonatal and paediatric practice could also contribute towards a better understanding infant neurological development.

Accelerometer and Sensor Data Methods

Motion sensors have seen extensive use in human motion analysis research as a means of detecting and tracking physical activity, particularly in healthcare related works. Physical motion sensors make use of a range of technologies including gyroscopes, accelerometers, and magnetometers,

Table 2.1: Overview of related works, methods used, features extracted, and reported accuracy

Method	Author	Features Extracted	Accuracy
Accelerometer	Heinze et al. [101]	Velocity, Acceleration	89%
Accelerometer	Singh et al. [102]	Acceleration	70-90%
Accelerometer	Gravem et al. [103]	Acceleration	70-90%
Accelerometer	Fan et al. [104]	Acceleration	82%
Electromagnetic Tracking	Karch et al. [105]	Temporal Stereotypy	n/a
Magnetic Tracking, Acc	Philippi et al. [106]	Kinematic Stereotypy	n/a
Accelerometer	Rahmati et al. [107]	Frequency	85%
Depth Sensor	Olsen et al. [108]	3D Reconstruction	n/a
Depth Sensor	Olsen et al. [109]	Velocity, Acceleration	92%
Accelerometer	Rahmati et al. [110]	Frequency	88%
CV/Accelerometer	Machireddy et al. [111]	Multi-modal	84%
Marker-based	Meinecke et al. [112]	Skewness, Periodicity	73%
Marker-based	Kanemaru et al. [113]	Vel, Acc, Periodicity	n/a
Frame Differencing	Adde et al. [114]	CoM, QoM, CPP	90%
Frame Differencing	Adde et al. [115]	CoM, QoM	88%
Optical Flow	Stahl et al. [116]	AMD, Freq, Wav	94%
Optical Flow	Rahmati et al. [110]	FFT Mean, FFT StD	92%
Optical Flow	Orlandi et al. [117]	Velocity-based	92%
Optical Flow	Raghuram et al. [118]	Velocity-based	66%
Optical Flow	Ihlen et al. [119]	CoM for 5s segments	87%
Optical Flow	Baccinelli et al. [120]	Vel, Acc, CoM	n/a
Deep Learning	Schmidt et al. [121]	Deep Features	65%
Deep Learning	Cunningham et al. [64]	SATCo Point features	92%
Deep Learning	Tsuji et al. [122]	Movement indices	90%

however, processing techniques have also been utilised to interpret visual data, such as depth sensor outputs, to reconstruct human motion. Each of these sensor types have seen prominent use in the functional observation of motor movements; for the assessment of neuro-muscular disorders (such as stroke and Parkinson's disease); for the evaluation of physical activities to identify disease patterns; for the quantification of rehabilitation therapy; and for the prevention of injury through motion analysis [123]. Recently the evaluation of infant movement has also been carried out using similar sensor data, to attempt to establish the viability of using these methods for the prediction of CP. These approaches typically analyse infant movements based upon the criteria specified in the physical examinations discussed in Section 2.1.1, here we discuss several of the most prominent sensor-based methods.

In [101], Heinze et al. used wired accelerometers to extract data for the production of a new feature set, which was based upon modelling the velocity and acceleration of infant movements. The dataset used consisted of motion data from 23 infants, composed of 19 healthy and 4 high-risk infants. A decision tree classifier was used and an overall accuracy of 89% was reported.

Singh et al. [102] also proposed a system which used accelerometers for the analysis of abnormal infant movements using a dataset consisting of 10 premature babies. Features based upon the maximum observed acceleration magnitude for the arms, the legs and the whole body were generated for classification, and accuracies of 70% to 90% were reported depending upon the relative cost of false positives and false negatives. However manual annotation of the data using additional accompanying videos was required to generate accurate labelled data. In an extension to this method, Gravem et al. [103] proposed a system consisting of five accelerometers which were placed on each of the infants' limbs and on their forehead. 10 infants were filmed for 1 hour, with the video data again used by assessors to provide annotation for the corresponding accelerometer motion data. The annotations were based upon the presence of Cramped Synchronised GMs, which were used to determine a binary classification of normal or abnormal. Once the motion data had been pre-processed, 166 features were extracted for classification. The study also reported a classification accuracy of between 70% and 90%, specifically in the detection of Cramped Synchronised GMs on this dataset. Fan et al [104] also extend this work by making use of this same dataset to assess statistical and temporal features for each measurement. They showed improved performance over the previous related method through the integration of AdaBoost and Naive Bayes

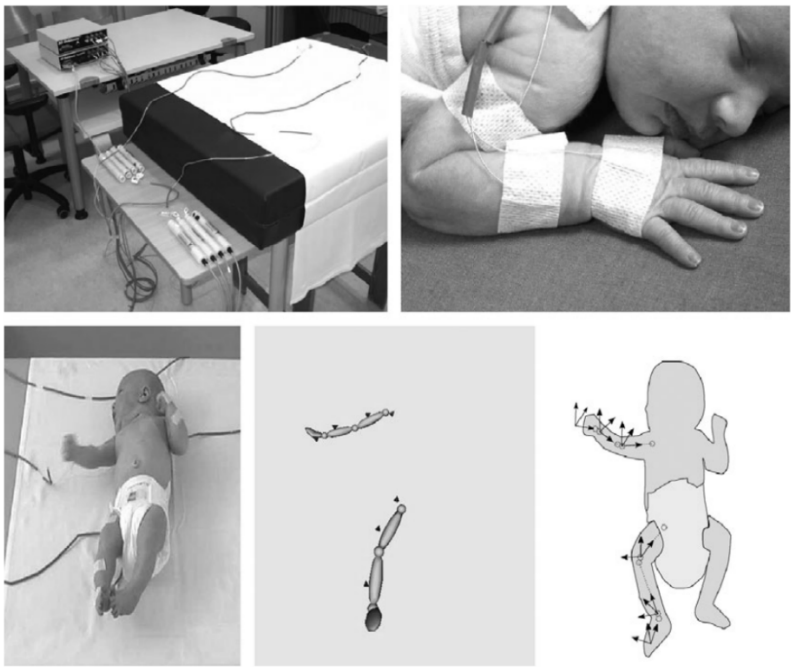


Figure 2.6: Example of kinematic recording using a magnetic tracking system [106]

classifiers. Additionally, the Erlang-Cox Dynamic Bayes Networks were also compared using Area-Under-the-Curve (AUC) for each model, with each providing comparable results.

Inspired by the successes in the previous methods, Karch et al. [105] attempted to model the limb movement of infants using an Electromagnetic Tracking System (EMTS). By attaching sensors to the limbs they calculated the joint centres based upon the Cartesian coordinates of the limb segments. In doing so they were able to establish the position and rotation of each joint centre from the captured motion data. After calculating the position and the rotation for each joint centre from the captured movement data, they assessed the recording accuracy by evaluating the differences between recorded and estimated sensor positions and orientations. Using a dataset consisting of 20 infant motions they suggest that they are able to demonstrate the segmental kinematics associated with abnormal infant movement using dynamic time warping. Using the stereotypy score and manual evaluation, they report a sensitivity of 90% and a specificity of 96%, but only using upper limb movements and on a small dataset. Additionally, distinctive individual motion features were not specified or quantified for this study and automated classification was not proposed.

In a subsequent study, Philippi et al. [106] used an accelerometer and magnetic tracking system to establish kinematic features to model repetitive movements in the limbs. They suggested that the repetitive movements shown in the upper limbs of infants in their dataset provided a predictive

indicator of later CP but did not propose an automated classification framework, instead providing a statistical analysis of the feature stereotypy as a means of manually predicting later CP based upon graphical representations of the associated feature data.

Given the success in the predictive capability of limb movements through the analysis of accelerometer data, Rahmati et al. [107] also attempted to model these movements using motion sensors, also making use of accompanying RGB video camera data. They carried out motion segmentation and extracted three features: area out of standard deviation from moving average; periodicity; and correlation between trajectories. These features were then used to classify the sensor data using the SVM classifier. Following on from this, in [110] they carried out frequency-based analysis on the accelerometer data. The accelerometers were again attached to the limbs of the infant and a set of features were extracted which reflected the frequency component of the movements in an effort to model the repetitive nature of the motion for further classification. They carried out a comparison with video-based features and reported that the results from the sensor data were not as accurate as those of the computer-vision features.

In an effort to bridge the gap between motion sensors and computer-vision, advanced range imaging techniques have subsequently been implemented, which use specific sensor technologies to produce depth images. These RGB-D cameras, such as the Microsoft Kinect, provide estimated 3D joint positions using a random forest [124], which can be further enhanced by introducing human prior knowledge [125], allowing them to be used effectively for motion monitoring [126].

In [108] and [109], Olsen et al. proposed the use of the Microsoft Kinect depth sensor to generate RGB-D imagery of infants for motion analysis. This RGB-D video data was used to construct a simplified 3D model of the infant, based on simple geometric shapes and a hierarchical model, for tracking. Using a dataset of 7 RGB-D videos of infants, they manually annotated frames to serve as ground truth and used this to evaluate the accuracy of their 3D model by calculating the euclidean distance between each. This model was then used to compute several features based on the angular velocities and acceleration to detect spontaneous infants movements for classification using annotated data to identify infants at risk of CP. However, whilst depth sensors offer a high temporal resolution, privacy preservation, and the potential for comprehensive analysis, the implementation of such as system can also pose significant practical problems, as they require complex

setup, can be difficult to use with infants, require dedicated equipment, and are often unsuitable for clinical use. This is also true of electromagnetic tracking systems and wearable sensors (accelerometers and IMUs) which are generally expensive, large, logistically difficult to utilise with infants, and are impractical for both in clinic and remote analysis; as such researchers started to explore purely image-based techniques in an effort to tackle these challenges.

Marker-based Methods

Marker-based motion capture is an image-based technique which is used to track movement using recognisable markers which are attached to an object or person, allowing for proprietary systems to interpret the associated motion. These systems are commonly used for human motion analysis by placing markers at import parts of the body, such as joints, and tracking spatio-temporal variations in body-part movement. This form of motion capture has also been widely used commercially and has recently been implemented to assess infant movement.

One of the first marker-based systems proposed for the detection of CP was the study by Meinecke et al. [112]. In this study, 7 infrared cameras fed by 20 reflective markers (Figure 2.7) were used to capture the 3D motion of 22 infants to assess the associated spontaneous motor activity. Quantitative features were then extracted from the motion data, and a combination of 8 features were used to differentiate between normal and abnormal infant movements based upon data annotations provided by GMA assessors. They reported a classification accuracy of 73% using a quadratic discriminant analysis algorithm.

A similar approach was used by Kanemaru et al. [113], where reflective markers were tracked and different kinematic features established. The aim in this study was to determine if the movements of infants who were later diagnosed with CP were 'jerkier' than those of healthy ones. However the main concern with using these techniques is that a sufficient number of markers/cameras must be used for reconstruction, and self occlusion can therefore cause significant problems. To address this Machireddy et al. [111], proposed a hybrid system which combined inertial measurement units (IMU) with marker-based video imagery. The aim here was to combine these techniques to overcome the shortcomings of each by producing a synchronised multi-modal dataset. They report a classification accuracy of 84% in identifying fidgety movements using a SVM classifier on a dataset consisting of 20 infants.

Figure 2.7: Motion analysis system, measurement setup and marker placement [112]

However, motion capture systems such as these are typically costly, cumbersome, complex to install, require significant fine-tuning, and have high computational complexity. Moreover, similar to the the use of sensors, to extract the spontaneous movement data from infants, markers may hinder or influence their movements, subsequently affecting the value of the data collected. Whilst these systems can produce reasonable results at modelling human motion, these detrimental factors are exacerbated by use on infants, making them largely impractical and subsequently limiting their viability for infant motion analysis in clinical environments [127]. To address this, marker-free approaches have been developed which make use of advanced computer-vision techniques to extract motion data from 2D RGB video. This approach offers cost-effective, non-invasive, non-intrusive alternatives, which address many of the challenges mentioned. Additionally, given the proclivity of 2D RGB video-capture equipment, it is suggested that this approach could much more easily be implemented at a significantly larger scale, supporting the proposed suggestion of viable infant screening programmes.

Frame Differencing Methods

One of the earliest examples of computer vision-based automated prediction of CP was the preliminary study by Adde et al.[128], who proposed a system in which "frame differencing" techniques were employed to assess the feasibility of undertaking automated GMA. In this study, infants were

videoed from above whilst lying in a supine position per GMA guidelines. Using this 2D RGB video as input, a representation of the associated motion was created by removing the background and calculating the difference between two consecutive frames in the video sequence. A point value per pixel of 0 or 1 was then assigned to represent the presence of movement, generating a "motion image". These motion images, which represented the amount of infant movement present in each video frame, were then used to generate relevant features for binary classification of normal or abnormal infant movements.

Figure 2.8: Frame differencing example. In the image on the right, pixels displayed in black indicate no movement occurred between the frames, pixels displayed in white represent movement [128]

This frame differencing method was evaluated in several subsequent studies to establish the viability of automated assessment using this approach, and to develop a feature set for further classification [114, 115, 129]. The engineered features consist several variants based upon the Centroid of Motion (CoM), and the Quantity of Motion (QoM), where the CoM is the spatial centre point of the motion image and the QoM is the sum of the associated motion of the centroid (further details of these extracted features are discussed in Section 5.1)

However the use of frame differencing images is subject to several limitations, such as sensitivity to camera movement, issues with self-occlusion, strict background pre-processing, and a lack of information about the speed and direction of objects moving in frame [116, 130]. As such, several studies have subsequently attempted to further evaluate the viability of automated examination through more advanced computer vision-based techniques, such as optical flow.

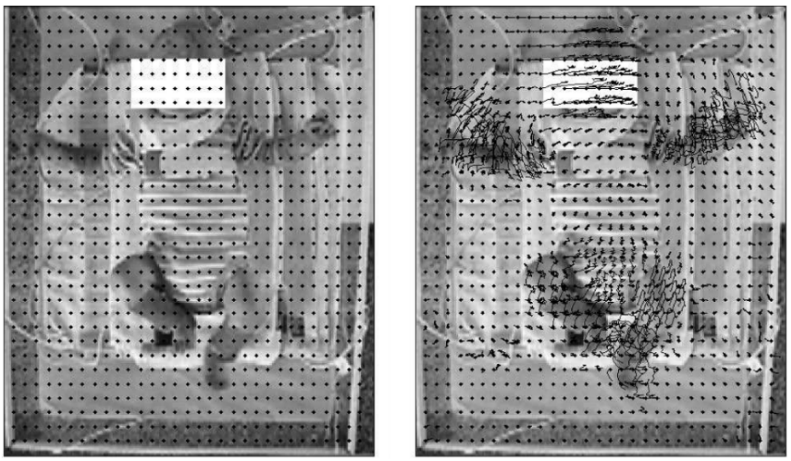


Figure 2.9: Example of motion trajectories obtained by tracked grid points using optical flow [116]

Optical Flow Methods

Optical flow is a method by which the motion intensities of an image may be calculated, as such the motion of objects within a scene can subsequently be detected and tracked across adjacent frames. Optical flow has been shown to be sensitive enough to be used to detect and delineate objects which are independent of one another, allowing for additional contextual information to be generated [131].

In optical flow based methods, the motion is represented by a vector field showing the displacement of the pixels between consecutive frames of an image sequence; rather than simply calculating the gross quantity of motion present in the frame and localising this centrally, as previously proposed in the frame differencing methods [116]. This allows for more detailed tracking of infant body-parts and the associated movements, subsequently providing opportunities for deeper analysis of the correlated motion. This approach is also better able to deal with camera motion through improved contextual understanding of the image composition, something which is particularly troublesome for the frame differencing methods discussed.

In the study by Stahl et al. [116], an optical flow-based method was produced which predicted cerebral palsy based upon statistical analysis and pattern recognition of the infant's spontaneous movements. Wavelet frequency analysis was used to evaluate the time-dependant trajectory signals found in the optical flow data. However, using this optical flow method presented issues with tracking larger movements, and as such it was suggested that video captured at a higher frame rate would be required for future analysis.

In order to address this problem, several works implemented Large Displacement Optical Flow (LDOF) to track infant body movements and extract features for classification [110, 117, 118, 119]. The LDOF method extends optical flow to better deal with large displacements of foreground objects and camera movement [132], making it more suitable for detailed infant motion analysis. Inspired by this and by previous sensor-based methods, Rahmati et al. [110] used LDOF to generate new motion features. In this case the use of features derived from frequency analysis to classify infant motion into one of two groups was proposed. Using video sequences as input, a motion segmentation algorithm was used to extract motion data from each limb. Their proposed feature selection method, which determined features with significant predictive ability, was then used prior to classification.

Similarly, Orlandi et al. [117] used LDOF to calculate the displacement of each pixel every 10 frames, in order to track movements and obtain the associated velocities. Background subtraction was also implemented, and features were extracted and classified as typical or atypical based upon the GMA, using several classification algorithms. Following on from this Raghuram et al. [118], used the same LDOF feature extraction techniques and motor parameters. In this case a skin model was also applied to the the infant silhouette in each video frame and the velocity was extracted using only pixels associated with the silhouette. These features were then used as a basis for predicting and screening for motor impairment.

The study by Ihlen et al. [119], also proposed the use of an LDOF model to track infant movements through a pixelwise representation. However in this case, short 5-second long segments were used as input with each annotated for the presence of risk-related movements characterised as normal or abnormal. The centroid of motion, rather than the centre of mass or anticipated joint position, of these tracked movements was then fed into a classification pipeline to determine the likelihood of cerebral palsy based upon the proportion of CP risk-related movements, allowing for statistical analysis of the data, rather than direct comparison with a predefined set of rules governing the classification. However this method attempts to model the proportion of risk related movements, and as such requires extensive labelled training data to determine the risk associated with each 5-second movement. Additionally holistic classification is compared with manual assessment rather than comparable automated models, and the use of CoM limits the interpretability and possibility of more comprehensive analysis.

The proposed methods suggest that the enhancement provided by LDOF allows for improvements to existing motion quantity features, the development of velocity-based features, and the introduction of frequency-based features for infant motion analysis and binary classification. Through comprehensive analysis of these movement-based features, it was subsequently determined that dynamic features are typically more predictive than the previous statistical features [110]. However, similar challenges are still present, such as issues with occlusion, drift, noise and computationally complexity [133], as well as susceptibility to unrelated movements (such as equipment, parents or clinicians in shot), and sensitivity to illumination changes [130]. Additionally, the low generalisability and difficulty providing suitable interpretability of extracted features make translation to clinical practice less viable [130, 110].

Deep Learning Methods

Recently the field of deep learning has provided state-of-the-art performance in many classification tasks. Deep learning is seen as advantageous as it is able to automatically learn features from input data without the need for feature engineering or specific domain knowledge. Deep learning can extract high level representations from complex data to produce a desired output classification and learn correlations and dynamic relationships that can be difficult to model through engineered features. As such, deep learning has seen extensive use in the field of Human Action Recognition, with its popularity continuing to grow in this area of research [134]. There have therefore been efforts to make use of deep learning frameworks for infant motion analysis [135].

In an initial study, Schmidt et al. [121] propose the use of a recurrent deep neural network to analyse a video dataset to classify infant fidgety movements. This study made use of transfer learning using the Keras VGG19 model, trained on the 1,000 classes of the ImageNet database. Image features were extracted from the framework and passed through an LSTM layer for classification. This resulted in a classification accuracy of 65% using a 10-fold cross validation strategy with an 80% training 20% test/holdout data split. They suggest that these results, whilst comparatively poor, are not unexpected as the project is in the early stages and with further work a more robust result is expected.

In a similar project, Cunningham et al. [64] used deep learning to undertake the clinical SATCo examination. In this work 13 keypoint features were identified and used to estimate the location

and orientation of the head, multi-segmented trunk and arms from videos of the clinical test. Experts manually annotated the 13 keypoint features for the image frames used and a CNN was then used to generate keypoints for unseen data. The generated data was then used to calculate segment angles for classification based upon the SATCo testing criteria. However, whilst a high accuracy of 92% was reported, poorer precision (84.2%) and F-Score (76.0%) were also reported, with these being more insightful metrics given the imbalance in the associated dataset.

In the paper by Tsuji et al. [122] a markerless movement measurement and evaluation system was proposed for the identification of GMs in infants. The proposed system uses an artificial neural network with a stochastic structure to calculate 25 movement indices related to GMs for further analysis. A log-linearized Gaussian mixture network (LLGMN) is then used as the classifier for the movements, reporting a classification accuracy of 90.2%. Whilst this method makes use of deep learning for classification, the manually extracted features are of low detail, since a similar method to that of the CoM is used. In future works they suggest that their proposed method could be further improved by the inclusion of additional detail in the feature extraction process, such as through the pose-estimation methods we propose.

2.3.2 Summary

In this section, we have discussed several automated methods for the prediction of CP, based upon the physical examinations discussed in Section 2.1.1. We discussed that whilst accelerometer, sensor-based, and marker-based methods can provide notable predictive accuracy, the logistics of implementing such approaches make clinical adoption impractical. As such, computer vision-based approaches have been explored, and are now largely considered to be the preferred method of conducting such assessments.

The computer vision-based approaches discussed range in complexity from the initial frame differencing methods, which provide a simplified interpretation of infant movement, through to optical flow based methods which are capable of extracting significantly greater levels of motion detail, and finally through to deep learning based methods which typically represent the state-of-the-art in image analysis. The evolution of these methods has largely been driven by challenges specific to each approach.

In the case of frame differencing, it was determined that there was a lack of information relating

to the speed and direction of objects in frame for suitably reliable prediction, as well as technical limitations, such as sensitivity to camera movement and issues with self-occlusion.

In the case of optical flow, difficulties with tracking larger movements became apparent, as well as issues with self occlusion, sensitivity to lighting changes, lack of contextual awareness (i.e. the motion data does not easily assess body parts separately, typically relying upon whole body movement analysis), and difficulty dealing with external influences (such as the parent or assessor being present in frame).

In the case of deep learning, whilst excellent performance is theoretically possible, the main challenges relate to the amount of data required for this predictive task and the lack of interpretability. Typically, deep learning methods rely upon significantly more data than traditional machine learning frameworks to extract meaningful features, and this is often not available in this particular domain. Additionally, the extraction of these deep features means that clinicians are largely unable to extrapolate the reasons why a deep learning framework might arrive at a classification decision, which, whilst acceptable in standard computer-vision tasks, is not suitable for medical assessment. Based upon these observations, we suggest that new methods are required, which make use of advances in the fields of computer-vision and human activity recognition, yet retain interpretability and robustness to the specified challenges, such as pose-estimation, body-part segmentation, and histogram representation techniques.

2.4 Relevant Techniques

2.4.1 Pose-Estimation

Inspired by the results in human motion analysis using accelerometer and marker-based methods, computer-vision researchers designed systems capable of estimating human skeletal pose from RGB video. These pose estimation techniques make use of advances in computer-vision and machine learning, to detect human figures in images and video, and determine their pose by estimating the spatial locations of key body joints.

The automated estimation of human pose from 2D images is an active research area, with several significant recent contributions [15, 16, 17, 18]. With the continued progression in deep learning techniques, various robust frameworks have been proposed which can accurately estimate human

Figure 2.10: Example pose estimation results obtained using OpenPose [17].

poses from 2D images. One of the most widely known methods is the work by Cao et al. [17], who present a framework, known as OpenPose, which can accurately detect the 2D pose of people from a single RGB image. The OpenPose framework provides excellent pose estimation performance from 2D RGB video input, producing an output consisting of the Cartesian coordinates of joint positions, based upon pre-defined keypoints, allowing for an output which incorporates both the joint position and orientation of the associated connecting bones (Figure 2.10). This allows for localised analysis of the motion data, subsequently addressing many of the challenges faced by other methods in the literature. The use of 2D RGB video data is particularly pertinent to our use case as it encourages accessibility through a reduced requirement for any specialist equipment, and interpretability through the use of human recognisable data which is also inherently anonymised.

2.4.2 Body-part Segmentation

The aim of human body-part segmentation is to partition individuals within an image into multiple semantically consistent regions (head, arms, legs, etc) [136]. This process is considered vital for many human-centric analysis applications [137]. Several recent studies [138, 139, 140, 141] have proposed methods to improve body-part segmentation through the combined training of human body-part segmentation and human pose estimation frameworks. However, these approaches are largely successful due to supervised training using pixel-wise manual labelling. Given that this is impractical for the large datasets required to train such a system, other methods have been developed to aid with this labelling requirement.

In the paper by Fang et al. [142], a semi-supervised approach is proposed, which uses data augmentation to transfer the human-labelled part segmentation from an existing dataset to a separate unlabelled dataset. In the work by Bearman et al. [143], a point-level supervision is used to create object priors which generate foreground masks. Whilst these approaches provides good

Figure 2.11: Example segmentation results obtained using [25].

performance in extracting foreground information, it has difficulty in determining the boundaries between foreground objects, this being particularly true of self occlusion. Given the nature of the data analysed in our project, the capability to adequately deal with self occlusion is important in this task. To address these issues, a recent work on multi-person part segmentation [25] proposed cross-domain complementary learning (CDCL), which provides state-of-the-art performance in human body-part segmentation tasks without the need for human-annotated segmentation labels. By using synthetic data to learn the boundaries between different parts, the CDCL system is also better able to deal with self occlusion. In order to extract meaningful body-part information from an input image, the CDCL pre-trained body segmentation model produces 6 contextual segments, or image masks, relating to the head, torso, upper arms, lower arms, pelvis and upper legs, and lower legs which can be used for both tracking body-parts and visualisation. An example of the segmentation output result is illustrated in Figure 2.11.

2.4.3 Histograms for Human Action Recognition

Histogram-based approaches have seen wide use in several fields for a number of years, and have been found to perform well in visual recognition tasks. Histogram-based approaches, such as

[144], have been successfully implemented in human action recognition tasks by condensing data into a lower dimensional range whilst also retaining the most useful information, providing a full but manageable impression of the associated data. Furthermore, the combined use of different kinds of histogram features has been shown to improve the accuracy of action recognition frameworks significantly [145, 146, 147].

A recent successful and relevant implementation of these techniques can be found in the work by Xia et al. [148]. In this paper, a histogram-based method was used in conjunction with joint positions extracted from Microsoft Kinect RGB-D depth data, to undertake classification of human actions into one of ten indoor activities. This approach allowed for the generation of feature descriptors, which were used to examine the distribution of both the orientation and displacement of each of the joints over a period of time. In doing so, this method bypassed the need to solve the time misalignment and variations in speed between two frames, as well as offering a robust solution to differing video durations. This approach makes it particularly well suited to the analysis of infant motion data due to the variability of the associated input video, further encouraging accessibility throughout the pipeline.

2.4.4 Summary

In this section, we have discussed that due to the limitations inherent in the related works, a new approach is required that takes advantage of advances in human motion analysis. As such, we examined pose estimation techniques and discussed how this approach would be advantageous over previous methods, particularly in addressing factors the previously discussed methods had difficulty dealing with, such as camera movement, external influences, varying illumination, resolution inconsistencies, differing subject scales, and larger body-part displacements. We explored the practical aspects of using pose estimation, with regards to retaining interpretability and allowing for data sharing through the built-in anonymity of the extracted pose information. We examined how image segmentation techniques might be used to provide additional contextual feedback and visualisation, affording greater interpretability to the classification process. Finally, we discussed that by employing histogram-based methods we can extract meaningful information from the raw data, whilst reducing dimensionality, and retaining human explainable feature representations which can be engineered to reflect specific assessment criteria.

Chapter 3

Methodology

In this chapter we provide an overview of the relevant methods that we employ within our proposed frameworks and how these might fit together to form a feature extraction, diagnostic classification, and visualisation pipeline. Our proposed frameworks seek to address several of the issues highlighted in the review of the related works, and as such we also discuss our intention for full re-implementation of several methods from the literature for comparative testing across multiple datasets, including videos captured in a real-world clinical setting.

3.1 Overview of the Proposed Methodology

In this thesis we propose and evaluate several new pose-based features which are directly mapped to motion criteria set out in the associated physical clinical assessments. We generate a real-world dataset, to inform and evaluate each of our proposed features, and use these features to verify the predictive value of our proposed computer-vision-based frameworks for the prediction of CP. Importantly, given that this study is based within the medical domain, we also incorporate interpretability into all stages of the proposed pipeline, such that the feature extraction, generation and classification frameworks remain transparent and explainable throughout. Additionally, one of the added benefits of our proposed pose-based method is that the extracted pose-data is inherently anonymised, and as such we are able to share the labelled pose dataset with the community.

The research aims in this thesis can therefore be summarised as:

- A pose-based method which can suitably represent infant motion to facilitate classification and subsequent prediction of CP.
- The collection of a real-world dataset to inform our feature design, and for evaluation of our proposed frameworks and baselines from the literature.
- Pose-based features which are mapped directly to the motion characteristics of clinical assessments for classification with suitable accuracy for clinical adoption.
- Interpretability incorporated into the proposed features and classification frameworks.
- A series of features which are robust enough to generalise well across multiple datasets and achieve state-of-the-art performance.

An overview of our proposed general pipeline is provided in Fig 3.1. In this pipeline we make use of 2D RGB video sequences as input. When it comes to capturing infants movements, we suggest that the use of 2D RGB video is preferred to that of RGB-D for two main reasons. Firstly, RGB video is much more accessible as it requires no specialist equipment (a camera-phone is sufficient), and secondly, RGB-D requires the emission of infrared light, which may have a health impact upon infants. The first stage in our pipeline is to extract pose data, consisting of 2D joint positions, from the 2D video sequences using the OpenPose framework. We then automatically pre-process the extracted pose data to ensure compatibility and consistency throughout the re-

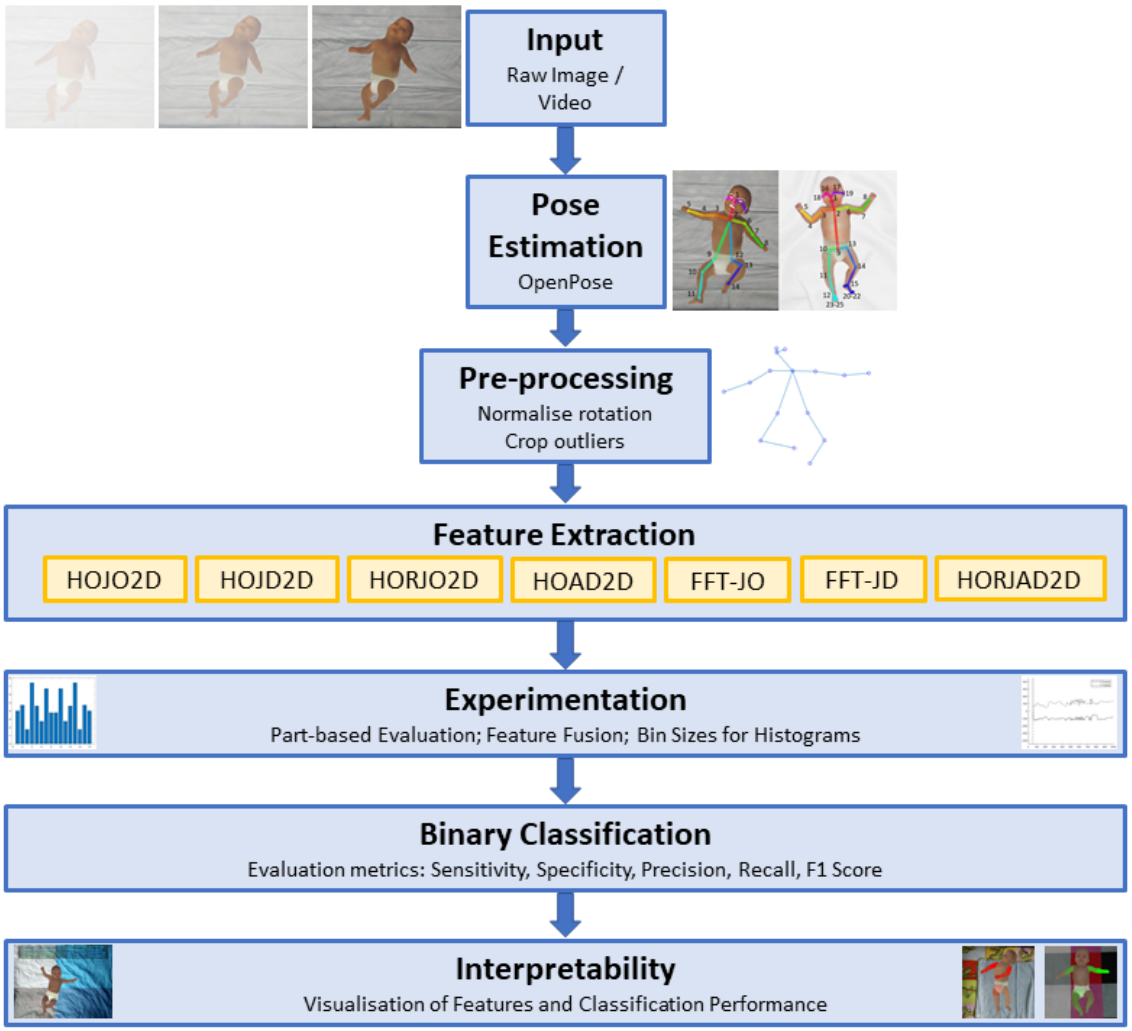


Figure 3.1: The overview of the proposed prediction and visualization framework.

mainder of the pipeline. Inspired by the work in human action recognition by Xia et al. [148], we propose techniques to produce histogram representations of infant movements from the processed data to generate motion features for binary classification. We then generate visualisations of the classification decisions with a view to highlighting motion abnormalities, such that assessors are better able to interpret the data thus improving model explainability. As such, we propose that through our use of pose-based assessment we aim to extract meaningful, human understandable data from standard 2D RGB video data, which is more robust and better able to deal with real-world clinical situations than existing methods.

3.1.1 GMA Informed Motion Features

To accurately analyse infant movements for the subsequent prediction of CP, suitable features must be generated for the classification algorithms to make use of. These features need to fully represent the motion characteristics of the dataset used, and as such the quality of the features has a significant impact upon any insights provided by the machine learning framework.

To generate features which are not only informative, discriminative, and robust, but also directly mapped to the assessment criteria in the clinical guidelines, we employ feature engineering techniques to generate features based upon domain knowledge provided by clinicians. Given that GM assessors typically look for specific movement patterns, we attempt to model these patterns through a set of orientation-based, displacement-based, and frequency-based features. Specifically we aim to model the movements associated with the assessment criteria set out in the GMA checklist [47], and the passive movement assessment section of the Optimality Score neurological examination [42]. Namely, we extract the following features, which we discuss in more detail in Section 5.2:

- Histograms of Joint Orientation (HOJO2D)
- Histograms of Joint Displacement (HOJD2D)
- Angular Displacement (HOAD2D)
- Relative Joint Orientation (HORJO2D)
- Relative Joint Angular Displacement (HORJAD2D)
- Fast Fourier Transform of Joint Displacement (FFT-JD)
- Fast Fourier Transform of Joint Orientation (FFT-JO)

Each of these features provides a histogram-based representation, comprised of an n-dimensional vector, which is used as the input for classification. These new histogram-based features therefore represent various motion characteristics extracted from 2D skeletal poses, specifically for infant motion analysis related to physical clinical examinations. By approaching this problem in this way we are able to flexibly assess infant motion both temporally and spatially, by looking at the distribution of the orientation, displacement, and frequency components of the joints over a specified period of time. We are also able to analyse part-specific motion for greater insight into

the movement associated with physical examination; and through temporal segmentation we can both explore the sequential aspect of these assessments, and also holistically examine the overall movement structure. The histograms generated also retain an element of human interpretability, allowing clinicians to intuit differences in distributions that align with manual observations found in the clinical physical assessments, as discussed in Section 6.2.5.

In addition to individual feature analysis we fuse complementary features to improve the overall performance and robustness of the extracted features across multiple datasets. By concatenating different features as early fusion, it is expected that better classification performance can be achieved if the classifier is capable of handling and learning the input data with higher dimensionality when compared with the individual features. The case-specific methods used for feature fusion, and the progressive classification evolution of our extracted features are discussed in greater detail in Chapter 6.

3.1.2 Binary Classification for CP Prediction

Once we have extracted the features, we then feed them into our machine learning classification framework to obtain an overall analysis of the movement characteristics and subsequent prediction of CP based upon the annotations provided by the GMA assessors. These annotations are based upon the observed motion characteristics derived from the physical examinations discussed in Section 2.1.1. The classification framework categorises the data into classes labelled as 'normal' or 'abnormal' based upon the annotations and the automated observation of the patterns in the associated feature data. In our works we experiment with both traditional machine learning frameworks and deep learning frameworks for this classification task, full details of which are provided in Chapter 6. Our aim here being to improve upon the diagnostic performance of other reported methods to ensure clinical viability whilst utilising relevant, human understandable features.

3.1.3 Baseline Re-implementation for Comparative Evaluation

Another issue found in the many of the existing methods in the literature is a lack of direct comparison in classification performance between different approaches. Whilst reasonable results have been reported in the related works, evaluation of the robustness of the proposed methods has not

been adequately demonstrated. Moreover, due to the sensitive nature of the video data recorded, the datasets used in the previous works are not available to the public. As a result, it is difficult for researchers to have a fair comparison and evaluation on the performance of different methods due to the unavailability of benchmark datasets. In our work, we re-implement several prominent CP prediction frameworks and compare the classification performance on different datasets. Details of the datasets used in our evaluations are discussed in Section 4.2 and Section 4.3, details of the baselines used in our comparisons are provided in Section 5.1, and full comparative evaluation the frameworks are discussed in Chapter 6.

3.1.4 Visualisation for Interpretability

One of the main issues with using machine-learning approaches for healthcare related tasks is the problem of interpretable AI. Whilst machine learning-based frameworks have obtained excellent performance in a wide range of visual understanding tasks, most of the existing frameworks are considered 'black-box' approaches, since most classification frameworks only output the predicted label without specifying exactly what influences the classification decision. As such, it is becoming increasingly important for the frameworks to ensure that the decision making process is transparent and fully understandable [149].

In order to make our proposed frameworks more interpretable, we include automatically generated visualization modules to aid the user/assessor. These modules highlight and provide additional pertinent information relating to the body-parts which are showing movement abnormalities. We are able to do this through the part-based assessment provided by our pose-estimation based method and through the use of GMA relevant features, transparent classification methods, and data visualisation techniques, meaning that interpretability is present throughout all stages of our pipeline. Details of the visualisations produced in our works and the associated quantitative and qualitative evaluations are provided in Chapter 7.

Chapter 4

Data Collection and Pre-processing

In this chapter we discuss the data collection methods used, the study design and ethical approval details, the composition of the datasets gathered, and the pose estimation and pre-processing undertaken on the datasets prior to further analysis.

4.1 Study Design and Ethical Approvals

For this collaborative project, a retrospective cohort study design was implemented for data collection, to produce the RVI-GMA dataset (detailed in Section 4.3). Full ethical approval was obtained from the host organisation, the Research Ethics Committee (REC), the Health Research Authority (HRA), and Health and Care Research Wales (HCRW) with the following reference numbers and dates of approval:

- Northumbria University Ethics Online: Ref:9865, approval granted 04/10/2018.
- Caldicott Approval: ID:6935, approval granted 17/12/2018.
- Northumbria University internal IRAS review: Research and Innovation Services approval granted 01/03/2019.
- REC: Ref:19/LO/0606, IRAS project ID: 252317, favourable opinion granted 21/05/2019.
- HRA and HCRW: REC Ref:19/LO/0606, IRAS project ID: 252317, approval granted 21/05/2019.

A study specific consent form, parent information sheet, and research protocol were produced (refer to Appendix A) and approved prior to data collection, so that fully informed consent could be obtained from the parents/legal guardians of all participating infants for the use of video recordings. Written informed consent was obtained by the clinical staff associated with the project prior to data collection. The study population included infants who had a clinical GMA, with a video recording at 3–5 months post-term, as part of their routine follow-up care.

The Moving INfants In RGB-D dataset (detailed in Section 4.2) is an open source dataset which was also used throughout the development of our project. All terms specified in the author's license agreement [150] were met.

4.2 Moving INfants In RGBD Dataset

One of the challenges facing researchers attempting to automate the GMA is the availability of suitable data. Given that the video data required for the GMA is of a sensitive nature, baseline datasets are not currently publicly available. Additionally, since human pose estimation frameworks are almost exclusively trained and tested using images of adults, a dataset consisting of

images of infants for research purposes can understandably be difficult to obtain.

To address this problem the Moving INfants In RGB-D (MINI-RGBD) dataset was generated and made publicly available to the community [150]. This dataset maps real-world 3D infant movements, captured in a clinical setting, to virtual 3D models of infants. Photo-realistic videos of the 3D infant models were produced using computer graphics rendering (640x480 resolution), allowing for the generation of anonymised, and subsequently shareable footage, which retains the real-world movement characteristics required for the GMA. This dataset consists of 12 top down videos of infants lying in a supine position, each 40 seconds in duration (1000 frames at 25 frames per second). We make use of this synthetic dataset and label each of the video sequences as 'normal', where fidgety movements are present (FM+), or 'abnormal', where fidgety movements are absent (FM-), based upon the GMA. The data labelling was carried out by two assessors highly experienced in the clinical application of the GMA, resulting in 8 videos being annotated as FM+ and 4 videos being annotated as FM-.

4.3 Royal Victoria Infirmary - General Movements Assessment Dataset

An important part of this study is that the framework has to have the ability to generalise well across different datasets, particularly when processing real-world video data. To reflect this, the challenging new Royal Victoria Infirmary - General Movements Assessment (RVI-GMA) dataset was collected to inform the design of our framework, and for evaluative analysis. This dataset is composed of real patient video data gathered as part of routine clinical care at the NHS Royal Victoria Infirmary, Newcastle upon Tyne, UK. As such, this dataset reflects the genuine intra-class variance and subsequent complexity present in the real-world clinical setting. The RVI-GMA dataset currently consists of 38 videos, of 38 different infants aged between 3 and 5 months post-term.

The videos were recorded using a standard 2D RGB handheld video camera (Sony DSC-RX100 Advanced Compact Premium Camera with 1.0-Type Sensor, 28-100 mm F1.8-4.9 Zeiss Lens, recording with a resolution of 1920x1080 @ 25 FPS). The duration of each video varied between a minimum of 40 seconds and a maximum of 5 minutes, with an average duration of 3 minutes and 36 seconds. The footage was captured from above, in a top-down orientation, with the infant lying

in a supine position, during active wakefulness, per the GMA guidelines. However, unlike many of the related works (e.g [128, 115, 24, 119, 117]), all video recordings were used as part of our evaluation, with no prior screening for inconsistencies such as poor lighting, camera movement, external factors, or significant shadows. Whilst this produced a more challenging dataset, it also represents a real-world evaluation of footage captured in a clinical setting. By including this challenging footage we hope to demonstrate that our proposed systems are more capable of being robust to variations in data capture, making them more suitable for clinical implementation. The videos were classified by two experienced assessors, using the GMA, into one of two categories; 1) FM+ where the infant demonstrates normal movements indicative of typical development; and 2) FM- where the infant demonstrates abnormal movement patterns that may be of concern to clinicians. This resulted in 32 videos being annotated as FM+ and 6 videos being annotated as FM-. In our published works two variants of this dataset were used, the first was the RVI-25 dataset which consists of the first 25 videos transferred from the Royal Victoria Infirmary (19 videos annotated as FM+ and 6 videos annotated as FM-), and the RVI-38 dataset which consists of all 38 videos. We implemented two variants of the dataset due to logistical constraints and as such were able to publish papers using both versions stated.

4.4 Pose Estimation and Data Pre-processing

Whilst the infant movement data is captured as raw RGB video, directly analyzing these videos is a challenging task since a wide range of factors contributes towards the intra-class variations, including: illumination, the background of the video, the appearance of the infant (body shape, skin color, with or without clothing), and external influences such as parental or clinician intervention etc. To address these issues we use the raw video as input for a pose estimation framework to compute joint positions, as discussed in Section 4.4.1. The extracted pose data is then corrected to remove outliers and inconsistencies, as discussed in Section 4.4.2. This corrected data is then used to generate features based upon the GMA for further analysis, details of the features used and the extraction processes are discussed in Section 5.2. These extracted features are then fed into classification frameworks, consisting of several different classification algorithms, for evaluation as discussed in Chapter 6 and Chapter 7.

4.4.1 Pose Estimation from Video

With the advancement of convolutional neural networks (CNNs), high performance can be achieved in solving a wide range of visual recognition tasks such as object detection, image segmentation and pose estimation. For pose estimation, the locations of body parts (e.g. joints) can be detected from an image. In particular, OpenPose [151] is one of the top-performing approaches proposed in recent years. OpenPose is based on Part Affinity Fields, which learn the association between body parts and their appearance in the image. Such an approach is also referred to as a 'bottom-up' approach that recognizes lower level features (e.g. body parts) first, in order to reconstruct the higher level skeletal posture.

18 Keypoint Pose Estimation

In this method, the official OpenPose implementation (https://github.com/CMU-Perceptual-Computing-Lab/openpose) is used for extracting the 2D locations of the joints from the video. Specifically, each video is converted into a sequence of images and a skeletal pose is extracted from each image. For each posture, 18 keypoints including body joint locations and facial landmarks are detected. An example of the skeletal pose extracted using OpenPose is shown in Figure 4.1. Each keypoint contains the x and y coordinates of the joint location within the image. In this method, 14 joints, consisting of *head, neck, left and right shoulders, left and right elbows, left and right wrists, left and right hips, left and right knees, and left and right ankles* were used.

25 Keypoint Pose Estimation

In order to extract further meaningful features for subsequent analysis and classification, we again make use of the OpenPose framework [151] to extract joint positions from 2D RGB video data. This time the extracted joint positions form a skeletal pose representation consisting of 25 predefined keypoints, as shown in Figure 4.1. We implement the updated 25 joint version of OpenPose since the authors suggest that this modified version provides improved accuracy. As such, each frame of each video is represented by 25 sets of 2D (x and y) coordinates and an associated confidence score for the prediction. In this method, we make use of all the extracted joints with the exception of the facial landmarks (joints 16 to 19), and the feet (joints 20 to 25), as it was determined that these joints were less reliable than the other body landmarks acquired through

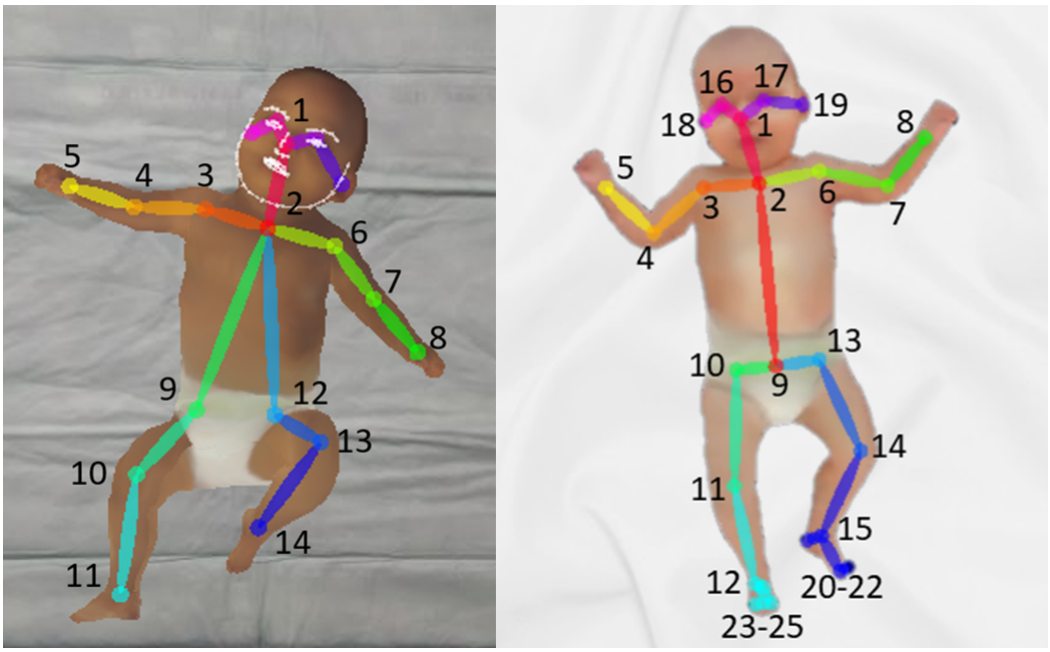


Figure 4.1: Examples of the extracted 18 keypoint and 25 keypoint pose estimation using OpenPose [151], with associated joint reference numbers overlaid on an example input RGB image from the MINI-RGBD dataset [150].

OpenPose due to occlusion/framing errors, and were found to play a less important role in the final GMA-based classification results.

4.4.2 Automatic Data Correction

To ensure consistency throughout the pipeline, the exported OpenPose data is pre-processed prior to feature extraction. This is because the accuracy of joint location prediction can be affected by factors such as self-occlusion of body parts. To alleviate this problem, an automatic data correction pre-processing approach is used. This pre-processing involves remapping anomalous joint positions caused by self occlusion or inaccuracies in the OpenPose joint prediction process. An example of this is shown in Figure 4.2. Given that OpenPose returns a confidence score associated with each predicted joint location, the first stage is a qualitative evaluation of the extracted OpenPose data, to check that predicted joint positions and the associated confidence scores correctly align, and are consistent with the input video. We then use the predicted confidence scores to calculate a confidence threshold. Since different confidence score distributions are obtained from different videos, due to the different movements as well as environmental conditions (such as lighting, video quality, etc), we adaptively adjust the threshold of the confidence score to decide

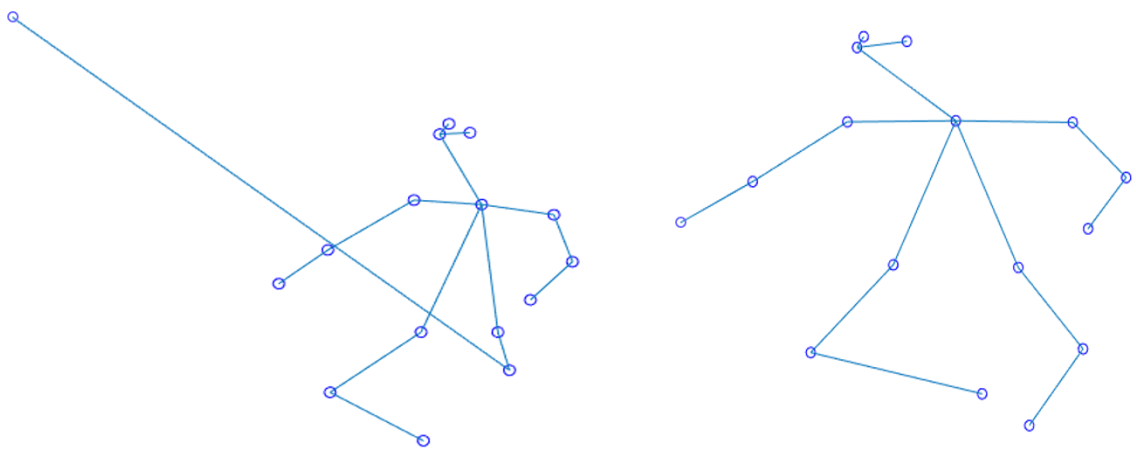


Figure 4.2: An example of our proposed automated pose data correction approach

whether the predicted joint location is 'usable' or correction is required. As such, the confidence threshold is calculated by taking the average confidence score, per joint, across each video sequence, and subtracting 5% from this. This means that we are able to remove joint positions with a lower confidence score than the confidence threshold, removing outliers on a frame by frame basis, for each joint, in each video sequence. As such, we compute the confidence threshold value t_i for joint i by

$$t_i = (\frac{1}{n}\sum_{j=1}^{n} c_{i,j}) \times 95\% \tag{4.1}$$

where n is the total number of frames (or postures), $c_{i,j}$ is the confidence score of joint i at frame j returned by OpenPose. From our experiments we empirically found that a manual threshold of 5% provided the best results.

Next, the trajectory of each joint is computed separately by curve fitting based on the joints with confidence scores which are above the threshold. In doing so, the location of the joint with a confidence score below the threshold at a frame will be estimated by the locations in the neighbouring frames with a higher confidence score. This aligns with the observation that human motion is continuous and the videos are captured at a high frame rate (25 FPS or above). As a result, the changes of the joint locations over time should be small and a curve function can approximate the joint trajectory over time. Among a wide range of curve fitting functions, the modified Akima interpolation [152] is selected in our work, since using this spline function can effectively avoid

the overshooting issues found in other spline functions. This results in a more natural interpolated trajectory, closer to the original signal. The joint locations with a confidence score above the threshold will then be used as the control point X_j as the input of the modified Akima interpolation.

$$X_j = [x_{j,0}, ..., x_{j,m}] \tag{4.2}$$

where m is the number of frames with a confidence score above the threshold, $x_{j,i}$ is the location of the joint j on the i-th 'usable' frames. The slope δ_i on interval between x_i and x_{i+1} can then be determined. The derivative d_i at the sample point x_i, which is used for the modified Akima interpolation, can be calculated by:

$$d_i = \frac{w_1}{w_1 + w_2}\delta_{i-1} + \frac{w_2}{w_1 + w_2}\delta_i \tag{4.3}$$

and the weights w_1 and w_2 are determined by:

$$w_1 = |\delta_{i+1} - \delta_i| + \frac{|\delta_{i+1} + \delta_i|}{2} \tag{4.4}$$

$$w_2 = |\delta_{i-1} - \delta_{i-2}| + \frac{|\delta_{i-1} + \delta_{i-2}|}{2} \tag{4.5}$$

.

We then applied a moving-average filter between frames on a per joint basis to reduce jitter present in the sequence. With the filter, smoother, more reliable movements were generated for motion analysis. Empirically, we found that using a filter calculated over a 5 frame sliding window provided the best results.

4.4.3 Data Normalization

The joint locations returned from OpenPose are presented as the x and y coordinates on the image. Since differences in body size may impact upon the magnitude of body movement, subsequently affecting classification accuracy, all of the extracted infant pose data is scaled to to a standardized height. We do this by computing the height of the infant from the skeletal pose. The height can

be estimated by the sum of the lengths of the following body segments: *lower leg, upper leg, hip-to-neck and head* (refer to Figure 4.1).

In addition to the scaling factor, the orientation of the infant also affects the performance of the machine learning process. As such, we transform the extracted pose sequence so that the infant pose data is vertically aligned and centred in each frame, as illustrated in Figure 4.2. We do this by computing the acute angle between the medial axis of the torso, which can be represented by a straight line between the middle of the hip joints (i.e. left hip and right hip for the 18 keypoint variant or joint 9 in the 25 keypoint variant) and the neck joint, and a vertical line in the coordinates system. Using the root joint as an example, we amend all joint coordinates so that the root is fixed at 0,0 whilst the relative distance of each of the joints is unchanged. We then calculate the rotation θ^f_{align} required at frame f to align the spinal column (i.e. the central line between joints 2 and 9) with the $y_{axis} = (0, 1)$ by

$$\theta^f_{align} = dir \times \arccos \frac{(fp^f_2 - fp^f_9) \cdot y_{axis}}{\|(fp^f_2 - fp^f_9)\| \|y_{axis}\|} \tag{4.6}$$

where fp^f_2 and fp^f_9 are the filtered 2D coordinates of joint 2 and 9, respectively, and $dir = sign(fp^f_2 - fp^f_9) \times y_{axis})$ is used to determine the direction of the rotation (i.e. clockwise or counter-clockwise). Finally, the normalized position p of each joint can be computed by

$$p^f_i = \begin{bmatrix} cos(\theta^f_{align}) & -sin(\theta^f_{align}) \\ sin(\theta^f_{align}) & cos(\theta^f_{align}) \end{bmatrix} (fp^f_i - fp^f_9)^T \tag{4.7}$$

where $i \in [1, 15]$.

This means that each posture in the motion sequence is rotated and centred according to the computed acute angle. We then apply these techniques to all of the extracted posture sequences. The pose-estimation methods utilised, the automatic data correction approaches used to deal with outlying anomalous data points, and the data normalisation techniques used ensure data consistency throughout, meaning they are normalized and ready for further analysis. With the data processed in this manner, we are then able to feed it directly into the next stage of our pipeline, namely feature extraction for classification.

Chapter 5

Feature Engineering

In this chapter we discuss several prominent methods from the literature, which are reimplemented and used as baselines for comparison. We also provide information relating to the feature engineering used to generate our proposed meaningful features, based upon the GMA, for the automated prediction of CP. Finally, we examine our proposed feature fusion methods and how these might further improve performance, and provide details of the required normalisation carried out prior to classification.

5.1 Baseline Features for Comparison

In order to assess the effectiveness and robustness of our system, we reimplement several video-based methods from the literature to serve as baselines for comparison, namely Centroid of Motion and Quantity of Motion [115] [129] [128]; Cerebral Palsy Predictor [114]; Absolute Motion Distance, Relative Frequency, and Magnitude of Wavelet Coefficients [116]; Frequency Analysis [110]; and the Movement Complexity Index [153]. We also compare our results with the methods reported in [154], [155] and [156] using the source code provided by the authors. From our literature review and the analogous survey paper by Irshad et al. [157], we suggest that these baseline methods represent the most prominent and comparable methods from the related works. To our knowledge, a comparison of the different proposed approaches has not been carried out to quantitatively evaluate the effectiveness of each method on shared datasets.

Centroid of Motion

In order to generate the Centroid of Motion (CoM), a motion image was first produced for each frame in the video sequence. A motion image is created by identifying the pixel variance between two sequential frames, and then assigning a pixel-wise point value of 0 or 1 based upon this variance. This process generates a frame-based image sequence, where white represents movement calculated between frames, and black represents no movement.

The CoM is the spatial centre point of the motion image which highlights the pixels with detected changes (i.e. body movement), and in our case, represents the the centre point of the movements of the infant. As discussed in the literature, the mean and standard deviation of the CoM in the x-and y-directions (CX_{mean}, CX_{SD}, CY_{mean} and CY_{SD}) were then calculated and exported as features for classification.

Quantity of Motion

The Quantity of Motion (QoM) is also calculated through the generation of a motion image. It is the sum of all pixels with positive values from the motion image, divided by the total number of pixels contained within the image. The standard deviation (Q_{SD}) and mean (Q_{mean}) of the QoM were calculated and used for classification. Both the CoM and the QoM were extracted using the musical gestures toolbox discussed in the literature (https://www.arj.no/2018/09/28/mgt/).

Cerebral Palsy Predictor

The Cerebral Palsy Predictor (CPP) [114] is the concatenated combination of the centroid of motion standard deviation (C_{SD}), the quantity of motion mean (Q_{mean}), and the quantity of motion standard deviation (Q_{SD}). Adde et al. [114] suggest that this feature combination is their best motion image based method of predicting of later CP. We concatenated the features as discussed and again carried out classification.

Absolute Motion Distance

The Absolute Motion Distance (AMD), Relative Frequency (RF), and Magnitude of Wavelet Coefficients (MWC) methods are all based upon optical flow information, which is used for motion-based tracking. For re-implementation, we followed the technical details presented in [116]. A regular grid was initialised and optical flow was used to track the motion of these grid points across a specified sequence of frames. Individual trajectories were therefore obtained for each of the tracked grid points allowing for further analysis of the motion.

The AMD is proposed as a holistic measure of activity, as it captures the absolute values of the optical flow velocities and stores them in histogram format. The histogram features are then classified. Since the bin size is not specified in [116], we empirically selected 64 as the size in our implementation. These histograms therefore represent the relevant statistical data about the amount of observed infant activity, and were used for classification.

Relative Frequency

Using the grid based optical flow data previously discussed, the Relative Frequency (RF) is generated by analysing the time dependent trajectory signals. The RF of the signal represents the occurring frequencies found in the movement patterns, this information is converted into a histogram for classification [116]. We followed [116] in our implementation. Again, the bin size of the histogram is not specified, so we use 64 in our work, as this was empirically found to provide the best results. In order to extract discriminative information relating the the frequencies, the time-distance between consecutive maxima is computed. A histogram representation of these measured time-distances is then generated and used as the relative frequency measurement and subsequently classified.

Magnitude of Wavelet Coefficients

The previously created optical flow trajectories were again used, this time to generate the Magnitude of Wavelet Coefficients (MWC). The wavelet power spectrum is used to demonstrate the variety of the observed movement at different resolution levels, providing insight into the complexity of the movement. We followed [116] to compute the histogram features from the wavelet power spectrum. The Meyer wavelet transform was used to convert the discrete sampled signal into a series of wavelet coefficients, which represent the amplitude of the wavelet function at a particular scale and location. A spectrogram of the trajectories was generated at specified power levels, computed as the the square sum of the detail coefficients of that level. The magnitude was then used as a measure of the movement activity and a histogram representation was once again generated and used for classification.

Frequency Analysis

Given that normal FMs are defined as an ongoing and variable stream of movements, Rahmati et al. [110] suggest that these motions can be better studied in frequency domain. As such, FFT was used to obtain the frequency components of the motion. We followed [110] to extract the mean and standard deviation values of the Fourier coefficients in horizontal (FFT-X_{mean} and FFT-X_{StD}) and vertical directions (FFT-Y_{mean} and FFT-Y_{StD}) as features for classification, using 100 bins with non-uniform sizes as specified in the literature.

Movement Complexity Index

Inspired by our work, Wu et al. [153], also make use of pose-based features for the prediction of CP using a feature extraction and classification pipeline. They proposed the Movement Complexity Index (MCI) which makes use of 2D keypoints extracted from RGB video alongside depth camera (RGB-D) footage to model infant movements in 3D space as a means of determining the complexity and correlation of the whole-body movement characteristics. This method primarily makes use of the joint angles as a means of extracting useful features using the Spearman Correlation Coefficient Matrix. However, this method focuses upon analyzing the features to predict the risk level of CP in the infant. The features are computed from 3D skeletal data which requires specialized image sensing devices to capture those data. Furthermore, the method requires the user to

specify a threshold level of the computed index to separate normal/abnormal, and determining this threshold automatically is not discussed. Whilst we report the result from this method in Sections 7.3 and 6.2.4, we are unable to fully verify the proposed framework on other datasets since it relies upon depth coordinate information provided by a Microsoft Kinect sensor or similar.

5.2 Proposed Features

In our work, we suggest that pose-based histogram features can effectively represent the motion and distribution of postures over time related to the GMA. Using the corrected and normalised pose data we therefore propose several new pose-based histogram features, for the analysis of infant body movements. The features presented here represent our continued refinement in modelling infant body movements and the evolution of our framework for the prediction of CP. As such, we provide details of the proposed features and their specific relevance to the motion characteristics and assessment criteria from the GMA.

Histograms of Joint Orientation (HOJO2D)

In this representation the 2D space is segmented into n bins that represent the prevailing angle of joint orientation. The joint orientation is computed by calculating the spatial relationship between a joint and its associated parent joint, subsequently allowing for the calculation of the alignment of the connecting bone:

$$bone = j_i - j_{i-parent}. \tag{5.1}$$

where j_i and $j_{i-parent}$ are the vectors containing the 2D coordinates of the i-th joint and its parent joint. We manually select the joint range (e.g. grouping joints 6, 7 and 8 to extract data pertaining to the left arm) to extract part specific information before a suitable bin is assigned for each joint per frame. As a result, the pose is represented by an n bin histogram of normalised data. Given that a key characteristic of abnormal infant movements is a lack of variability, this feature is therefore able to represent the uniformity of the infant pose, highlighting the prevalence of repetitive postures.

Histograms of Joint Displacement (HOJD2D)

In this representation the displacement of each joint (i.e. the euclidean distance travelled across a specified frame range) is extracted and recorded every 5 frames. Again a histogram-based approach is used to represent the displacements, with a relevant bin, each of which represents a regular incremental increase, being assigned based upon the degree of displacement. Again, a range of joints is selected manually for part-based analysis. In this way the displacement can be represented by an n bin histogram of normalised data. This feature is engineered to represent the quantity and variety of body-part specific movements exhibited by the infant.

Histograms of Angular Displacement (HOAD2D)

HOAD2D represents the change in angular orientation across a specified time interval for each body part in the video. This histogram-based feature captures the distribution of the angular displacement between a predefined regular offset interval, in which our experiments found 8 frames to be optimal. As such, the smoothness of the body part movements can be represented, for example, a smooth movement should be characterised by a histogram which has only a small number of bins having high values. This feature is therefore designed to help identify spasmodic, abrupt, and sporadic movements of short duration. The orientation of each joint is essentially the 2D vector pointing from the parent joint to the child joint:

$$o_i^f = p_i^f - p_j^f \tag{5.2}$$

where o_i^f is the orientation (2D vector) of joint i at frame f, p_i^f and p_j^f are the 2D coordinates of joint i and j, respectively, and joint j is the parent of joint i. Here, we further compute the angular displacement θ_i^f of joint i at frame f by:

$$\theta_i^f = \arccos \frac{o_i^f \cdot o_i^{f-\Delta t}}{\|o_i^f\| \|o_i^{f-\Delta t}\|} \tag{5.3}$$

where Δt is a predefined regular offset interval of 8 frames, the value of which we determined empirically from our experiments. Equation 5.3 is essentially the cosine similarity of o_i^f and $o_i^{f-\Delta t}$. By this, the direction (i.e. clockwise or counter-clockwise) of the angular displacement will be discarded such that the feature will solely focus on the magnitude of the orientation change.

Having computed the angular displacements of every joint in the whole pose sequence, the HOAD2D for each joint is computed by quantizing the displacements into a finite number of bins. Therefore, the number of bins and the size (i.e. range of values) of each bin will significantly affect the discriminative power of the feature. We observed that most of the angular displacements are very small, while the maximum theoretical displacement is 180°. As a result, we propose using non-uniform bin sizes to better represent the distribution for the angular displacements:

$$bs_i = \frac{180^\circ}{2^{n-i}} \tag{5.4}$$

where bs_i is the bin size for the i-th bin and n is the total number of bins for the histogram feature. For HOAD2D, we empirically found that $n = 16$ yields the best results.

Histograms of Relative Joint Orientation (HORJO2D)

To analyse the coordination and synchronisation of different body part movements, it is important to extract features from different joints simultaneously. Inspired by Rueangsirarak et al. [158], we propose representing the distribution of the relative orientation of the joints using a histogram-based feature. Here, the pairwise relative joint orientation is computed in a similar manner as in Equation 5.2:

$$o^f_{i \to j} = p^f_j - p^f_i \tag{5.5}$$

, although the two joints are not necessarily physically connected. In order to capture the synchronisation of different parts of the body, we compute the relative orientation for all pairs of joints.

Since the relative joint orientation with a range of 0° to 359° can be obtained, a uniform bin size is used and we empirically found $n = 16$ produced the best performance. Once the individual joint histograms have been extracted, we combine these to form histogram representations for each limb prior to concatenation for classification. HORJO2D intuitively represents the body synchronisation, as such, a histogram which has only a small number of bins having high values means that the joints are moving in the same direction together.

Histograms of Relative Joint Angular Displacement (HORJAD2D)

To further capture the change in body part movement synchronisation over time, the angular displacement of the relative joint orientation is also extracted as a histogram feature. This feature is crafted to evaluate the relationship between body parts, such that whole body coordination, dystonia, and ataxic movements can be assessed. As with extracting the HORJO2D feature, the pairwise relative joint orientation (RJO) is computed and similarly combined. We further compare the RJOs before and after the predefined frame offset interval, and angular displacement is calculated using the cosine similarity of the two RJO vectors similar to the calculation of HOAD2D, as in Equation 5.3.

Again, most of the angular displacements computed are having a small value. As a result, the non-uniform bin size (Equation 5.4) is used to increase the discriminative power of the HORJAD2D feature. From our experiments, we empirically found that $n = 8$ provides the best results.

Fast Fourier Transform of Joint Displacement (FFT-JD)

Whilst the aforementioned histogram features represent the distribution of different kinds of spatial features at a coarse level, the temporal ordering information is being discarded. Inspired by the previous work in analysing body movements in the frequency domain [110], we propose the FFT-JD feature. This feature contains the magnitude of each of the frequency components extracted from the motion such that the variability of the motion can be better assessed. By using the Fast Fourier Transform (FFT) we convert the extracted joint displacement signal $D_i = [\|\dot{p}_i^2\|, \|\dot{p}_i^3\|, \cdots, \|\dot{p}_i^m\|]$ of joint i from a motion with m frames to a representation in the frequency domain, allowing us to model the complexity, fluidity, and variability of the movements, whilst highlighting any repetitive, athetoid, tremulous, or myoclonic characteristics. Additionally it is reported that analyzing human motion in the frequency domain is more robust to noisy data [159], and as such helps with the task of assessing some of the smaller, more detailed movements associated with the GMA.

We extract the FFT-JD by applying FFT to the vector D_i:

$$Y_i^k = \frac{1}{m} \sum_{f=0}^{m-1} D_i e^{\frac{-l2\pi kf}{m}} \tag{5.6}$$

where Y_i^k is a vector which contains the magnitude of the frequency component at index k for joint i, $e^{\frac{l2\pi}{m}}$ is a primitive m^{th} root of 1.

Having computed the frequency component Y_i from D_i, Y_i is partitioned into 16 bins with non-uniform bin sizes:

$$bs_{FFT-JD,b} = \begin{cases} F\frac{b^2}{n^2}, & \text{if } b = 1, \\ F\frac{b^2}{n^2} - \sum_{k=1}^{b-1} bs_{FFT-JD,k}, & \text{if } 2 \leq b < n, \\ F - \sum_{k=1}^{b-1} bs_{FFT-JD,k}, & \text{if } b = n. \end{cases} \tag{5.7}$$

where $bs_{FFT-JD,b}$ is the size of the b-th bin and F is the number of frequency components obtained from D_i using FFT. The last bin (i.e. $bs_{FFT-JD,n}$) will occupy the remaining space.

Fast Fourier Transform of Joint Orientation (FFT-JO)

Similar to the FFT-JD feature we once again make use of FFT to model repetitive movements by looking into the frequency components. This feature provides information relating to the rigidity, directional variation, and range of movement associated with the infant's posture. In this case we model the repetition and frequency of similar postures from a joint orientation sequence $O_i = [o_i^1, o_i^2, \cdots, o_i^m]$ for joint i using FFT as in Equation 5.6. The histogram-based FFT-JO is computed using the same method described in Equation 5.7, where the frequency components are computed using O_i instead. In this case, the bins for the lower frequency components will be smaller and can more effectively represent the low frequency components, since GMs from infants are mostly in lower frequencies, while the high frequency motion signals likely contain noise, allowing us to use this feature to further remove potentially anomalous noisy data from the classification process.

Revised HOJO2D and HOJD2D

In our initial iteration of these features, we explained that HOJO2D and HOJD2D are the histogram representations of the infant pose based upon the orientation of selected joints and their displacements respectively. We further improve upon this previously reported method by extracting individual joint histograms and concatenate these to form limb-based representations. This method means that we are able to incorporate a greater range of motion data than the previous

method, which extracts an individual per-limb histogram representation grouping and amalgamating several joints together. This approach allows us to assess movements in finer detail such that we are able to gain a better understanding of the underlying motion characteristics.

Feature Fusion

In addition to evaluating the classification performance of each of the individual features, we also fuse our selected features together for further analysis. Our aim in applying feature fusion is that better classification performance will be achieved over using only the individual histogram features, and that better robustness across datasets will be evident. In our experiments we apply early fusion by concatenating relevant feature vectors creating a fused feature set for classification. We experiment on several combinations of fused features, with the results reported in Chapter 6 and Chapter 7.

To generate our first group of fused features we concatenate HOJO2D and HOJD2D. In this grouping, we focus on evaluating 3 features, consisting of Arms, Legs and Limbs, with both 8-bin and 16-bin variants of the individual features.

Our second group of fused features consists of *pose-based* features. This grouping represents the angular feature information extracted from the pose data, as such these representations are indicative of the overall quality of the infant posture and the predominant directions of movement. The *pose-based* features consist of a concatenation of HOJO2D, HOAD2D, HORJO2D, and FFT-JO.

Our third group of fused features consists of the *velocity-based* features. This feature set represents the displacement of the joints over predefined time intervals, and as such model the speed, fluidity, coordination, and complexity of the infant movements. The *velocity-based* features consist of a concatenation of HOJD2D, HORJAD2D, and FFT-JD.

Lastly, we fuse the *pose-based* feature set with the *velocity-based* feature set such that we can generate a holistic representation of the infant motion patterns.

5.3 Histogram Normalisation

With the features extracted, we need to ensure they are on the same scale prior to further analysis for consistency. In order to do this we use Z-score to standardise the feature data h, by

$$z = \frac{h - \mu}{\sigma} \tag{5.8}$$

where μ and σ are the mean and standard deviation of all samples in the training set, and z contains the normalized features. In our implementation we use Z-score as this allows our system to retain the shape properties of the original dataset, with our initial classification results showing improvements using this method over min-max normalization. We empirically found that this provided us with the most robust solution across all of the proposed methods and the different datasets, as such we employ this in each of the frameworks discussed in Chapter 6 and Chapter 7.

5.4 Concluding Remarks

In this chapter we have examined our proposed pose-based features. Specifically we have discussed the theory behind the feature extraction processes for each of our proposed features, namely: Histograms of Joint Orientation (HOJO2D), Histograms of Joint Displacement (HOJD2D), Histograms of Angular Displacement (HOAD2D), Histograms of Relative Joint Orientation (HORJO2D), Histograms of Relative Joint Angular Displacement (HORJAD2D), Fast Fourier Transform of Joint Orientation (FFT-JO), and Fast Fourier Transform of Joint Displacement (FFT-JD). We have also discussed the baseline features selected for comparison, namely: Centroid of Motion (CoM), Quantity of Motion (QoM), Cerebral Palsy Predictor (CPP), Absolute Motion Distance (AMD), Relative Frequency (RF), Magnitude of Wavelet Coefficients (MWC), Frequency Analysis, and Movement Complexity Index (MCI). These baselines represent the most prominent methods from the literature, and as such, an explanation of our implementation of each is provided. With the proposed features and baseline features extracted from the video data, we are able to feed each into our proposed machine learning frameworks, for classification.

Chapter 6

Machine Learning

In this chapter, we discuss our proposed machine learning frameworks for the prediction of CP. We provide details relating to the architecture of our frameworks, the evolution of our proposed diagnostic machine learning pipeline, and the metrics used to evaluate performance. We then provide further discussion concerning our evaluation of classification performance, feature robustness and generalisability across multiple datasets.

6.1 Proposed Machine Learning Frameworks

The overall motivation for our project is to evaluate the feasibility of using our proposed pose-based features, for the task of identifying CP in infants based upon the GMA and associated physical examinations. Our classification frameworks consist of several traditional machine learning algorithms and deep learning frameworks. Given that the optimal selection of a classification algorithm can be dependant upon several factors, we simultaneously assess both the proposed features and the selected classification algorithms. This approach allows us to generate an interpretation of the strength of the features assessed, and the relative performance of each classifier.

Each of the following sections represents a step forward in our overall framework's design. The framework discussed in Section 6.1.1 focuses upon the introduction and classification of our initial HOJO2D and HOJD2D features, and serves as a feasibility study into the viability of classification using pose-based assessment. In Section 6.1.2, we expand upon this and examine several deep learning architectures for classification of the HOJO2D and HOJD2D features, to establish the viability of a hybrid classification approach. In Section 6.1.3, we discuss the introduction of our proposed supplementary features, and how these fit into our improved classification framework. Finally, in Section 6.2 we provide extensive evaluation and discussion relating to classification performance of each of the discussed frameworks, and analysis of the effectiveness of the proposed features.

6.1.1 Initial Traditional Machine Learning Framework

In out initial feasibility study, the extracted pose data is based upon the 18 keypoint pose estimation variant of OpenPose as discussed in Section 4.4.1. We evaluate the effectiveness of using the the proposed pose-based features HOJO2D and HOJD2D, discussed in Section 5.2, for classifying video footage into two categories (normal and abnormal). In our experiments, the ground truth for comparison is the data annotation carried out by the independent expert reviewer using the clinical GMA.

To generate meaningful features, in this work, we manually select the joint range to extract part specific information before a suitable bin is assigned for each joint per frame. As a result, body-parts from each selected joint range are represented by an n bin histogram of normalised data.

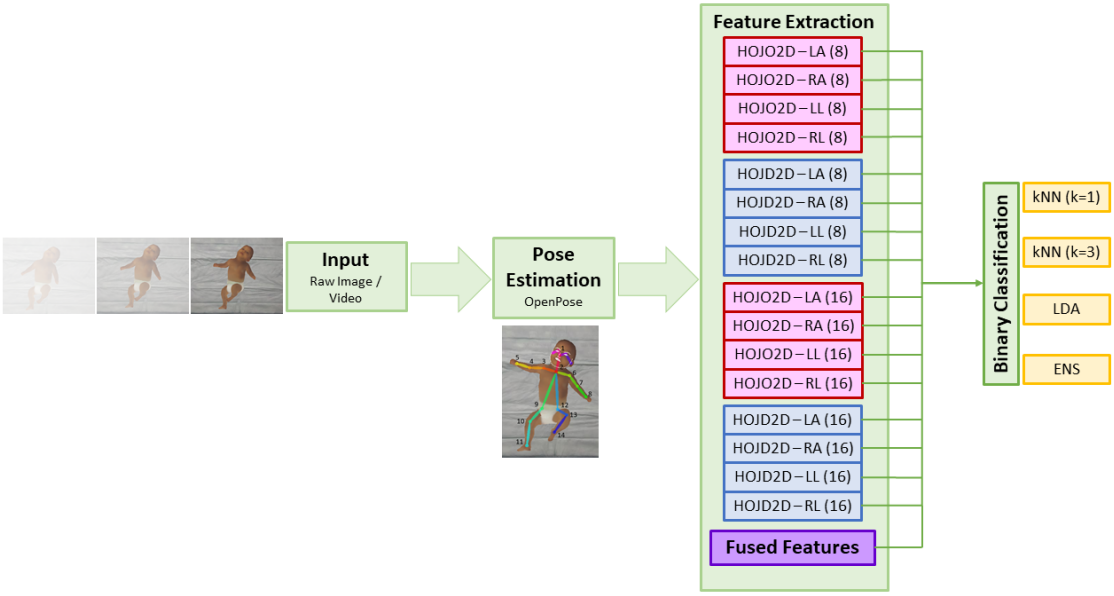


Figure 6.1: Our initial feature extraction and classification framework

This means that we are able to generate a part-specific histogram for each joint range selected, allowing us to feed these individually as inputs into the classification framework to get a part-specific classification result.

The proposed feature extraction and classification framework diagram (Fig 6.1) shows our extracted 8 and 16 bin features, in which LA refers to Left Arm, RA refers to Right Arm, LL refers to Left Leg, and RL refers to Right Leg. We train the k-Nearest Neighbour (kNN) (where k=1 and k=3), Linear Discriminant Analysis (LDA) and Ensemble (ENS) classification algorithms for binary classification, using different combinations of HOJO2D and HOJD2D, details of which are provided in the experimental results discussed in Section 6.2.2.

6.1.2 Proposed Deep Learning Architectures

Whilst the initial results in determining the feasibility of using our proposed HOJO2D and HOJD2D features using traditional classifiers were promising, we also wanted to evaluate the effectiveness of using a deep learning framework for this classification task. However, the holistic application of deep learning in the healthcare domain faces several challenges, most notably the large amount of data usually required for suitable results, and the problem of interpretable AI. Understanding how a framework arrives at a decision is particularly important in the healthcare domain, and this

is often very difficult, if not impossible to do with an end-to-end deep learning framework as deep features are typically incomprehensible for human perception. With this in mind, in this section we propose several deep learning frameworks which act simply to classify the hand-crafted features HOJO2D and HOJD2D, which we generate from the normalised, pre-processed data discussed in Section 4.4, using the methods discussed in Section 5.2.

The individual features as well as the fused feature sets are exported for evaluation in our classification experiments using five separate deep learning architectures. In particular, three distinct types of network architectures are proposed. We first introduce a fully connected network architecture which serves as a basic classification framework, we then further propose 1D and 2D convolutional neural network architectures. We carry out extensive experiments and ablation studies to determine the effectiveness of each proposed framework, details of which are discussed in Section 6.2.3.

The proposed deep neural network architectures are implemented in the PyTorch framework. We ran all experiments on a desktop computer with a single NVIDIA TITAN XP graphics card. Additional parameters such as $epochs = 4000$, $learningrate = 0.0005$ and $batchsize = 3$ were used in all tests.

Fully Connected Deep Networks

Fully connected deep network architectures are considered a generic framework for handling different problems since they are robust to different kinds of inputs (such as text, extracted features, images, videos, etc). Our proposed fully connected network architecture (Figure 6.2), namely *FC-Net*, is designed with gradually decreasing layer sizes. The input of the network is a 1D vector of the histogram-based features. The output of the last fully connected layer is fed into a softmax layer for classification. To reduce the negative impact of overfitting, we have constructed a system where each fully connected layer is followed by a dropout layer.

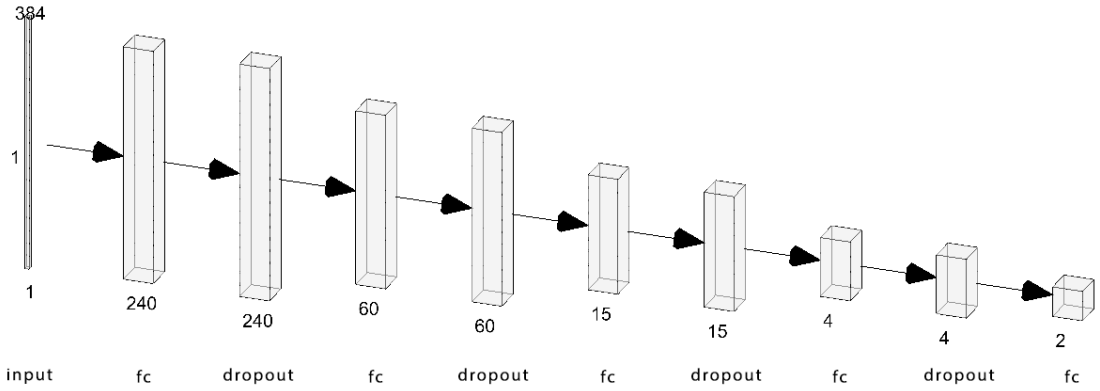


Figure 6.2: The proposed *FCNet* network architecture consisting of fully connected (fc) layers and dropout layers.

1D Convolutional Neural Networks

In the proposed pose-based features, the neighboring values are actually capturing similar body postures (i.e. with body part orientation in HOJO2D) and movements (i.e. with body part displacement in HOJD2D). In this framework, the limb-level and fused features are created by appending the histogram features of individual body parts resulting in a long 1D vector:

$$hist_{combined1D} = [hist_{part_1}, hist_{part_2}, ..., hist_{part_n}] \tag{6.1}$$

where $hist_{combined1D}$ is the final feature vector concatenated from the histogram features extracted from individual body parts and n is the number of body parts included in this feature.

To exploit the spatial information from the features, we propose two 1D convolutional neural network architectures, namely *Conv1DNet-1* (Figure 6.3) and *Conv1DNet-2* (Figure 6.4), to attempt to learn any deep representations and subsequently improve performance.

Due to the relatively low dimensionality of the input feature vector, both of the proposed architectures contain two 1D convolution layers. To further improve the performance, each 1D convolution layer is followed by a max pooling layer to down-sample the output, further feeding into a dropout layer to avoid overfitting. Similar to *FCNet*, the input of our network is a 1D vector of the histogram-based features. The output of the last dropout layer is flattened into a 1D vector and the dimensionality is reduced by a fully connected layer before feeding into a softmax layer for classification.

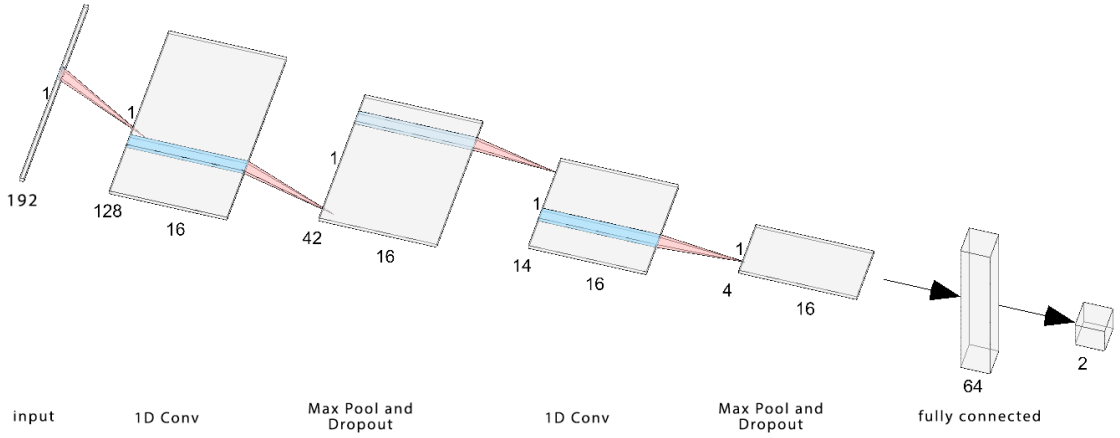


Figure 6.3: The proposed *Conv1DNet-1* network architecture which consists of 1D convolution, max pooling and dropout layers.

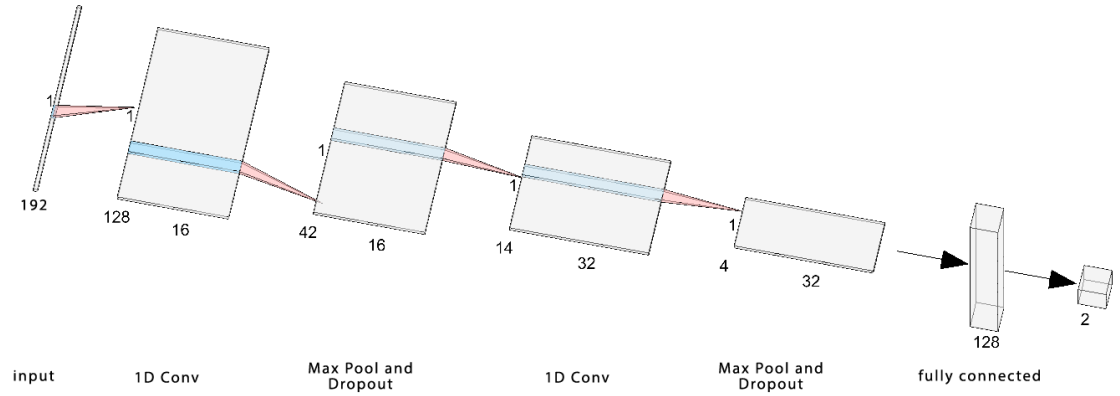


Figure 6.4: The proposed *Conv1DNet-2* network architecture which consists of 1D convolution, max pooling and dropout layers. Note the gradually increasing output channel sizes.

Each of the convolutional layers use the same group of settings, with $kernel\_size = 3$ and $stride = 3$. For the max pooling layers, $kernel\_size = 3$ and $stride = 3$ are also used. The main difference between the two networks is that *Conv1DNet-1* (Figure 6.3) has a constant output channel size while *Conv1DNet-2* (Figure 6.4) increases the output channel sizes gradually.

2D Convolutional Neural Networks

To further exploit the spatial information among different body parts in the motion, we further propose two 2D convolutional neural network architectures. To learn the spatial correlation within the 2D convolutional neural network, the input vector has to be converted into a 2D matrix shape. This is done by reshaping the 1D feature vector to a 2D matrix with each row containing the

histogram features extracted from a single body part:

$$hist_{combined2D} = \begin{bmatrix} hist_{part_1} \\ hist_{part_2} \\ \vdots \\ hist_{part_n} \end{bmatrix} \tag{6.2}$$

The two 2D convolutional neural network architectures we propose, namely *Conv2DNet-1* (Figure 6.5) and *Conv2DNet-2* (Figure 6.6), share a common design, with two 2D convolution layers. Similar to the proposed 1D convolutional neural networks, each 2D convolution layer is followed by a max pooling layer to down-sample the output and further feed into a dropout layer to avoid overfitting. The output of the last dropout layer is flattened into a 1D vector and the dimensionality is reduced by a fully connected layer before feeding into a softmax layer for classification. All the convolutional layers are using the same group of settings with $kernel\_size = 3$ and $stride = 1$. For the max pooling layers, $kernel\_size = 3$ and $stride = 2$ are used.

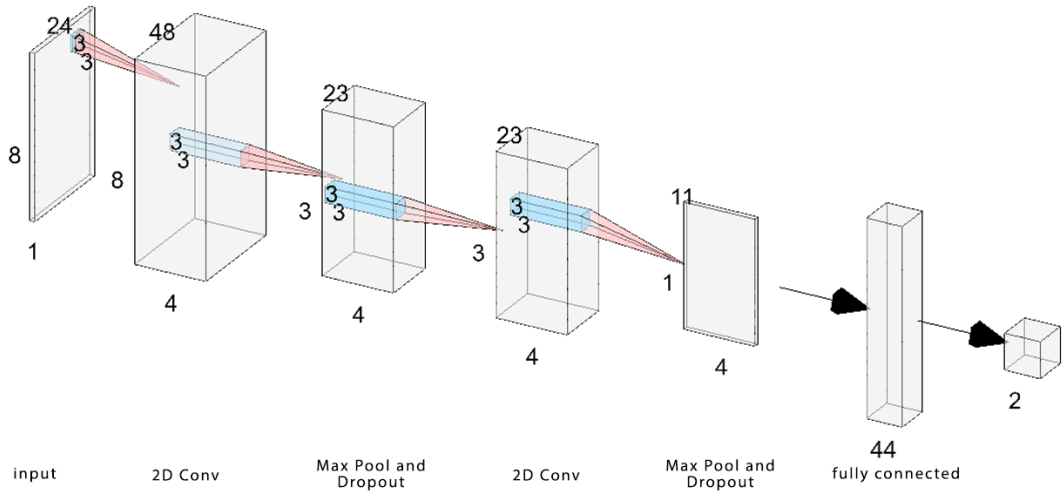


Figure 6.5: The proposed *Conv2DNet-1* network architecture which consists of 2D convolution, max pooling and dropout layers.

The main difference between the two networks is that *Conv2DNet-1* (Figure 6.5) has a constant output channel size while *Conv2DNet-2* (Figure 6.6) increases the output channel sizes gradually.

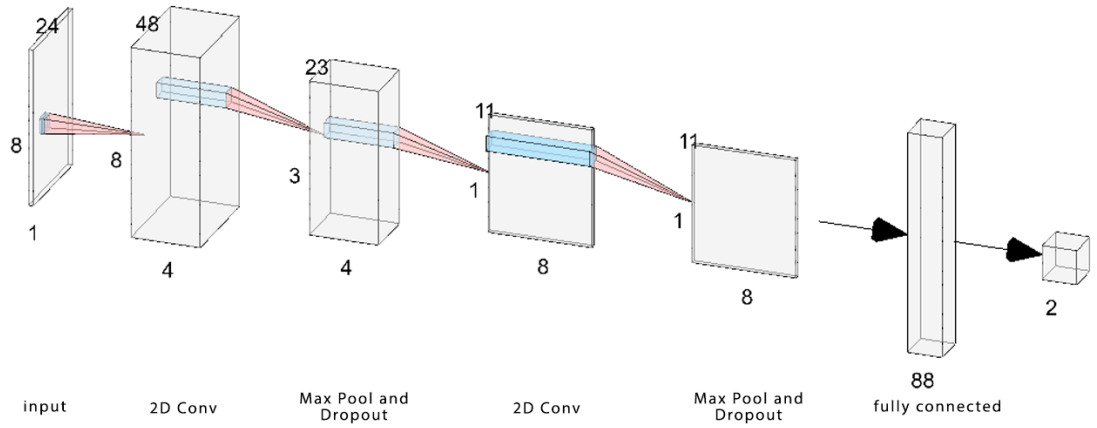


Figure 6.6: The proposed *Conv2DNet-2* network architecture which consists of 2D convolution, max pooling and dropout layers. Note the gradually increasing output channel sizes.

6.1.3 Improved Traditional Machine Learning Framework

In our previous frameworks we evaluated the feasibility of using the HOJO2D and HOJD2D pose-based features. In this section we provide details on how we integrate additional pose based features to enhance the predictive accuracy of our prediction pipeline. Given that GM assessors typically look for specific movement patterns, we attempt to model these patterns through a set of orientation-based, displacement-based and frequency-based features. As such, we propose several new features, specifically HOAD2D, HORJO2D, HORJAD2D, FFT-JD, FFT-JO, as well as the revised versions of HOJO2D, HOJD2D, as discussed in Section 5.2. These features are used as input to our improved machine learning pipeline (Figure 6.7) for further analysis.

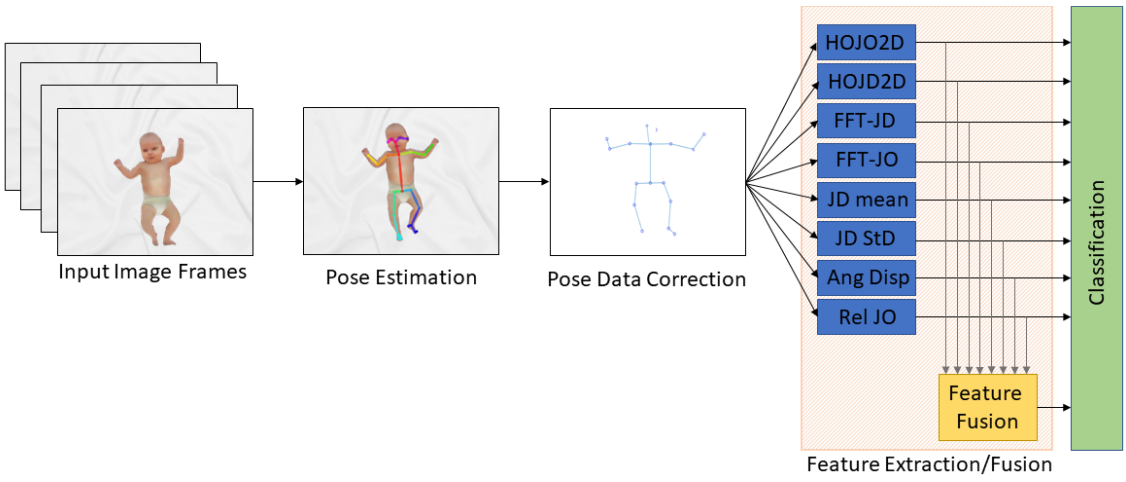


Figure 6.7: Overview of the pose estimation, feature extraction and classification framework.

In this proposed framework we again make use of traditional machine learning classification algorithms for practicality and interpretability. In this case, in addition to the classifiers used in our initial framework (Linear Discriminant Analysis (LDA), Ensemble (ENS), and k-Nearest Neighbour (kNN) where $k = 1$ and $k = 3$), we are using several other classification algorithms (specifically, Logistic Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT)), details of which are provided in Section 2.2.2.

We train the classification algorithms for binary classification, using different combinations of the extracted features, and evaluate the classification performance across multiple datasets to subsequently establish the feature robustness for this classification task. In order to do this we introduce a new real-world dataset which we use to inform and evaluate our newly proposed features. Finally, we implement several prominent baseline methods from the literature, which we use for comparison on shared datasets for an unbiased review of current assessment techniques. Full details of these experimental results are discussed in Section 6.2.2.

6.2 Experimental Results

In this section we evaluate the results of the various experiments on each of the proposed frameworks discussed in Section 6.1. We provide full details of the metrics, datasets, comparative baselines, relevant ablation tests, and feature analysis used in our evaluations. We also discuss our overall observations and assessment of the robustness of the proposed features and classification frameworks.

6.2.1 Performance Measures

In this project we make use of several evaluation metrics to assess the performance of each feature and the associated classifier; in this section we provide details of each of the metrics used.

$$AC = \frac{TP + TN}{TP + FN + TN + FP} \tag{6.3}$$

$$SE = \frac{TP}{TP + FN} \tag{6.4}$$

$$SP = \frac{TN}{TN + FP} \tag{6.5}$$

In our evaluation, True Positive (TP) is a measure of the cases in which impaired infants are correctly classified as impaired, True Negative (TN) represents unimpaired infants correctly classified as unimpaired, False Positive (FP) represents unimpaired infants incorrectly classified as impaired, and False Negative (FN) represents impaired infants incorrectly classified as unimpaired. Based upon these metrics, the Sensitivity (SE) is defined as the percentage of correctly identified positive classifications amongst the positive population of the dataset, the Specificity (SP) is defined as the percentage of correctly identified negative classifications amongst the negative population of the dataset, and the Accuracy (AC) is defined as the holistic percentage of all correctly classified instances.

$$PR = \frac{TP}{TP + FP} \tag{6.6}$$

$$RE = \frac{TP}{TP + FN} \tag{6.7}$$

$$F1 = 2 \cdot \frac{PR \cdot RE}{PR + RE} \tag{6.8}$$

We also calculate the Precision (PR) , Recall (RE) , and F1 Score (F1) , since accuracy metrics alone are generally considered to be insufficient to suitably determine the robustness of a classification model. PR represents the percentage of correctly identified positive cases from all of the positive predictions, RE measures the correctly identified positive cases from all of the actual positive cases, and F1 is the harmonic mean of PR and RE and as such conveys the balance between the precision and the recall.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{6.9}$$

Finally, we also calculate the Matthews Correlation Coefficient (MCC). The MCC is generally regarded as a particularly reliable statistical test since it only produces a high score if the prediction obtains good results in all of the four confusion matrix categories (i.e. TP, TN, FP, FN), proportionally both to the size of positive elements and the size of negative elements in the dataset. This characteristic makes it the preferred method of evaluating binary classification tasks on imbalanced datasets [160].

6.2.2 Initial Machine Learning Framework Classification Performance

In our evaluation of the initial framework discussed in Section 6.1.1 we make use of the MINI-RGBD dataset. Since this dataset has a limited number of video sequences (12 in total), we employ leave-one-out cross-validation to evaluate the performance. We train each classifier using different combinations of the HOJO2D and HOJD2D features extracted from 11 video clips and used these trained classifiers to predict the class label of the remaining unseen video clip. This process is repeated 12 times to make sure every video clip is evaluated. The average accuracy for each of the different feature combinations is reported in in Table 6.1, Table 6.2, and Table 6.3.

Evaluation of HOJO2D

In our experiments, we evaluate the HOJO2D feature at different levels based on the human body hierarchy. Specifically, we first extract the HOJO2D on each limb. Then, we apply feature fusion to the limb-based features to create a wide range of features: Arms – fusing features from both arms; Legs – fusing features from both legs; Limbs – fusing features from all 4 limbs. We also construct a single histogram, called Full body, by using the joint orientation from all body parts. Additionally, we further extract the features using both 8 bins and 16 bins, the results of which are presented in Table 6.1. The best classification accuracy (91.67%) is obtained using the 8 bin variants of the Arms and Limbs features, using the ENS classifier. In this setting, we generally find that individual features or fused features which contain the joint orientation computed from

Table 6.1: Classification accuracy using the HOJO2D feature

Features		Classification Accuracy (%)			
Type	Bins	kNN (k=1)	kNN (k=3)	LDA	ENS
Left Arm	8	50.00	66.67	66.67	83.33
Right Arm		58.33	16.67	25.00	25.00
Left Leg		33.33	66.67	58.33	33.33
Right Leg		50.00	50.00	33.33	83.33
Left Arm	16	75.00	75.00	83.33	83.33
Right Arm		50.00	16.67	33.33	33.33
Left Leg		16.67	41.67	58.33	3.33
Right Leg		33.33	41.67	66.67	33.33
Arms	8	75.00	75.00	66.67	**91.67**
Legs		25.00	41.67	50.00	41.67
Limbs		41.67	58.33	58.33	**91.67**
Full Body		66.67	50.00	50.00	58.33
Arms	16	66.67	66.67	75.00	66.67
Legs		25.00	8.33	41.67	33.33
Limbs		66.67	66.67	83.33	66.67
Full Body		83.33	50.00	66.67	75.00

the arms leads to better classification performance. In most cases, applying feature fusion achieves a better result than the basic individual part-based features. We also observe that, with this small dataset, features extracted using the 8-bin setting perform better than the 16-bin setting. This could be caused by the large pose variations, resulting in a diverged joint orientation distribution and therefore a reduction in the discriminative power as the number of bins increases.

Evaluation of HOJD2D

Again, we evaluate HOJD2D using features extracted from different limbs and apply feature fusion to generate new features for classification. The results are listed in Table 6.2. In this setting, the best classification accuracy (100.00%) is obtained with the 16 bin variants of the Right Leg and Legs features, using the ENS and LDA classifiers. Again, applying feature fusion generally achieves a better result than the basic limb-based features, which is consistent with the results obtained from classifying the HOJO2D features. In this case, features which contain joint displacement computed from the legs provide better classification performance. Also, the classification results obtained from features extracted using 16-bins perform better than the 8-bin setting. This indicates that for this feature set the 16-bin features are more discriminative and suggests that the magnitude of joint displacement is more consistent than the joint orientation.

Table 6.2: Classification accuracy using the HOJD2D feature

Features		Classification Accuracy (%)			
Feature	Bins	kNN (k=1)	kNN (k=3)	LDA	ENS
Left Arm	8	83.33	50.00	66.67	75.00
Right Arm		50.00	66.67	66.67	83.33
Left Leg		41.67	66.67	66.67	50.00
Right Leg		83.33	58.33	58.33	66.67
Left Arm	16	66.67	50.00	66.67	75.00
Right Arm		58.33	33.33	33.33	33.33
Left Leg		66.67	75.00	50.00	75.00
Right Leg		83.33	58.33	83.33	**100.00**
Arms	8	66.67	50.00	33.33	50.00
Legs		50.00	58.33	58.33	91.67
Limbs		75.00	75.00	75.00	91.67
Full Body		58.33	58.33	58.33	66.67
Arms	16	58.33	50.00	50.00	50.00
Legs		66.67	58.33	**100.00**	91.67
Limbs		58.33	58.33	83.33	91.67
Full Body		58.33	58.33	66.67	75.00

Table 6.3: Classification accuracy using the fused HOJO2D + HOJD2D features

Features		Classification Accuracy (%)			
Feature	Bins	kNN (k=1)	kNN (k=3)	LDA	ENS
Arms	8	75.00	50.00	66.67	**91.67**
Legs		50.00	58.33	50.00	**91.67**
Limbs		66.67	58.33	83.33	**91.67**
Arms	16	75.00	58.33	75.00	66.67
Legs		41.67	50.00	58.33	**91.67**
Limbs		66.67	66.67	83.33	66.67

Evaluation of Fusing HOJO2D and HOJD2D

In our next experiment, we evaluate early fusion of HOJO2D and HOJD2D. The results listed in Table 6.3 generally indicate that fusing HOJO2D and HOJD2D achieves better classification accuracy compared with using HOJO2D or HOJD2D individually. The best performance is obtained with all features with the 8-bin setting and the Legs feature with the 16-bin setting using the ENS classifier. This suggests that, through this consistency of performance, more robust classification is obtained by using the proposed fused features. Based upon these reported results, the fusion of both the part-based features and the different feature sets can therefore be considered to be the optimal grouping for future works.

6.2.3 Proposed Deep Learning Architectures Classification Performance

In our next series of experiments, we evaluate the proposed motion classification performance of the different deep neural network architectures discussed in Section 6.1.2. The MINI-RGBD dataset [161] is again used in all experiments. We compare the classification accuracy obtained from the proposed methods with baseline approaches and we employ a leave-one-out cross-validation approach, with the averaged classification accuracy again reported. We obtained the classification accuracy of all methods using three types of input features: 1) HOJO2D, 2) HOJD2D, and 3) fusing HOJO2D and HOJD2D (i.e. concatenating). Due to the random initialization of our newly proposed deep learning frameworks, the performance of the classifier may vary in different trials, as such we report the best performance of the classifiers in each setting. We also justify the selection of the hyper-parameters in the proposed network architectures by conducting a series of ablation studies to determine the effect the hyper-parameters have upon classification performance.

Performance using HOJO2D

The results using the HOJO2D features are presented in Table 6.4. In general, the newly proposed deep learning classification frameworks perform better than previous methods, as the majority of the highest accuracies (highlighted in bold) are obtained using our methods. In particular, *FCNet* performs well consistently achieving 83.33% across all of the different features. This highlights the generality of the fully connected neural network. The proposed *Conv1D-2* and *Conv2D-2* with gradually increasing output channel size in the convolutional layers also demonstrated high performance with most of the features having the same classification accuracy as *FCNet*. Accuracy obtained using *Conv1D-1* and *Conv2D-1* are lower than the other proposed frameworks, but they are more consistent and robust than the baseline approaches. For the baselines, the results are highly inconsistent. While some of the classification accuracies are high (such as the 8-bin Arms and 8-bin Limbs features with LDA), classifying some other features can result in very low accuracy (such as Legs with 16 bins). The results on this features set have subsequently demonstrated the high performance and robustness of the proposed deep learning frameworks.

Table 6.4: HOJO2D feature set: Classification accuracy comparison between our proposed deep learning methods and baseline machine learning methods

Histograms of Joint Orientation 2D (HOJO2D)							
Bins	8			16			
Features	Arms	Legs	Limbs	Arms	Legs	Limbs	Average
LDA	**100.00%**	75.00%	**100.00%**	75.00%	41.67%	75.00%	77.78%
SVM	66.67%	66.67%	66.67%	66.70%	66.70%	66.70%	66.70%
Tree	75.00%	0.00%	75.00%	75.00%	33.33%	75.00%	55.56%
kNN (k=1)	83.33%	25.00%	58.33%	**83.33%**	8.33%	33.33%	48.61%
kNN (k=3)	83.33%	25.00%	41.67%	**83.33%**	41.67%	58.33%	55.56%
Ensemble	75.00%	25.00%	75.00%	75.00%	8.33%	75.00%	55.56%
FCNet	83.33%	**83.33%**	83.33%	**83.33%**	**83.33%**	**83.33%**	**83.33%**
Conv1D-1	83.33%	75.00%	83.33%	**83.33%**	75.00%	75.00%	79.17%
Conv1D-2	83.33%	**83.33%**	75.00%	75.00%	**83.33%**	**83.33%**	80.55%
Conv2D-1	75.00%	**83.33%**	75.00%	**83.33%**	**83.33%**	75.00%	79.17%
Conv2D-2	83.33%	**83.33%**	83.33%	**83.33%**	75.00%	**83.33%**	81.94%

Performance using HOJD2D

The results using the HOJD2D features are presented in Table 6.5. Again, the newly proposed deep learning frameworks are more robust and performed more consistently. Whilst kNN (k=1) and LDA achieved some of the best accuracies with 91.67% on the 8-bin Arms feature and 100% on 16-bin Arms feature, the accuracy on other features are much lower (such as 50.00% and 41.67% on both Legs features). For our approaches, *FCNet* performed well and obtained 91.67% accuracy in Limbs features whilst achieving 83.33% in the rest of the features. The other deep learning frameworks are performing in a predictable manner, with variations in classification accuracy being between 75.00% and 91.67%, representing a smaller range than the other methods. This again highlights the robustness of the proposed deep learning frameworks.

Performance using Fused Features - HOJO2D + HOJD2D

The results using the fused features are presented in Table 6.6. Since the input feature size is doubled in this experiment, deep learning frameworks generally demonstrated a significant advantage in processing features due to this higher dimensionality. In particular, *Conv1D-2* achieved an excellent performance by having 91.67% classification accuracy in 5 out of 6 feature types. Whilst the LDA method obtained excellent classification accuracy on the 8-bin Arms and Limbs features, the results obtained in other features are significantly lower highlighting the inconsistency of the baseline methods, with accuracies ranging from 58.33% to 100%. *Conv1D-1* demonstrated a solid

Table 6.5: HOJD2D feature set: Classification accuracy comparison between our proposed deep learning methods and baseline machine learning methods

Histograms of Joint Displacement 2D (HOJD2D)							
Bins	8			16			
Features	Arms	Legs	Limbs	Arms	Legs	Limbs	Average
LDA	**91.67%**	41.67%	50.00%	**100.00%**	50.00%	75.00%	68.06%
SVM	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%
Tree	83.30%	50.00%	83.33%	66.67%	41.67%	66.67%	65.28%
kNN (k=1)	**91.67%**	50.00%	75.00%	**100.00%**	41.67%	75.00%	72.22%
kNN (k=3)	66.67%	41.67%	83.33%	58.33%	58.33%	66.67%	62.50%
Ensemble	83.33%	58.33%	83.33%	66.67%	58.33%	66.67%	69.44%
FCNet	83.33%	**83.33%**	**91.67%**	83.33%	**83.33%**	**91.67%**	**86.11%**
Conv1D-1	83.33%	75.00%	83.33%	83.33%	**83.33%**	83.33%	81.94%
Conv1D-2	83.33%	75.00%	83.33%	83.33%	**83.33%**	83.33%	81.94%
Conv2D-1	75.00%	**83.33%**	83.33%	91.67%	**83.33%**	75.00%	81.94%
Conv2D-2	83.33%	**83.33%**	83.33%	75.00%	75.00%	83.33%	80.55%

performance in achieving 91.67% using the 8-bin Legs feature and 83.33% using the remaining features. *FCNet* showed a robust performance again by obtaining 83.33% classification accuracy in all feature types. For the 2D convolutional neural networks *Conv2D-1* and *Conv2D-2*, the performance is once again consistent, with a small range of accuracy from 75.00% to 83.33%. We also observe that, in most cases, applying feature fusion achieved a better classification performance than the individual histogram features.

In summary, the experimental results on different feature types highlight the performance gain in both accuracy and robustness with the use of the proposed deep learning frameworks over the baseline approaches. The results also show that *FCNet* performed in a highly predictable manner with a relatively simple network architecture.

We also observe that, in general, the 16-bin variant is better for the proposed deep methods whilst the 8-bin version is better in the non-deep baseline methods. This is due to the fact that deep networks can better handle features in higher dimensionality than non-deep methods. This also suggests that the 16-bin features are more discriminative, particularly in the case of joint displacement, where the magnitude of the joint displacement appears to be more consistent for classification than the joint orientation.

We also note that when the dimiensionality of input features becomes higher, the benefits of using convolutional neural networks can be observed, as seen when using *Conv1D-1* and *Conv2D-1*

Table 6.6: Fusing the HOJO2D and HOJD2D feature sets: Classification accuracy comparison between our proposed deep learning methods and baseline machine learning methods.

Fused features - HOJO2D + HOJD2D							
Bins	8			16			
Features	Arms	Legs	Limbs	Arms	Legs	Limbs	Average
LDA	**100.00%**	66.67%	**100.00%**	**91.67%**	58.33%	83.33%	83.33%
SVM	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%
Decision Tree	75.00%	50.00%	75.00%	75.00%	25.00%	75.00%	62.50%
kNN (k=1)	91.67%	33.33%	83.33%	**91.67%**	50.00%	75.00%	70.83%
kNN (k=3)	91.67%	33.33%	83.33%	66.67%	58.33%	66.67%	66.67%
Ensemble	75.00%	58.33%	75.00%	75.00%	33.33%	75.00%	65.28%
FCNet	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
Conv1D-1	83.33%	**91.67%**	83.33%	83.33%	83.33%	83.33%	84.72%
Conv1D-2	83.33%	**91.67%**	91.67%	**91.67%**	**91.67%**	**91.67%**	**90.28%**
Conv2D-1	83.33%	83.33%	83.33%	75.00%	75.00%	75.00%	79.17%
Conv2D-2	83.33%	75.00%	83.33%	83.33%	75.00%	83.33%	80.55%

to classify fused features. This can be explained by the abstraction power of the convolutional layers in the network. We believe the performance gain of 2D convolutional networks will be even greater when the input features have even higher dimensionality (e.g. by incorporating time-series movement data).

Ablation studies

We conducted ablation studies to investigate the impact of the hyper-parameters on the classification performance. Since we have already compared the effect different layer sizes have on the proposed 1D (i.e. *Conv1D-1*, and *Conv1D-2*) and 2D (i.e. *Conv2D-1*, and *Conv2D-2*) network architectures, in this section, we focus on another hyper-parameter, namely the dropout rate. We picked the fused features setting in this ablation study, while training the networks with different dropout rates (i.e. 0.1, 0.3, 0.5, 0.7 and 0.9). The results are plotted in Figures 6.8 to 6.12.

The results show that the different dropout settings result in similar classification accuracy. Typically, the best performance occurs when the dropout rate equals 0.5 or 0.7, while some good performance can be obtained when the dropout rate equals 0.3. For the more extreme values we see a drop in accuracy, with 0.1 being unlikely to produce the best performance, and 0.9 producing the worst performance. Whilst there are some variations in the classification accuracy using dropout rate settings, the range is relatively small when compared with the inconsistent performance from baseline approaches, showing less sensitivity to hyper-parameters changes.

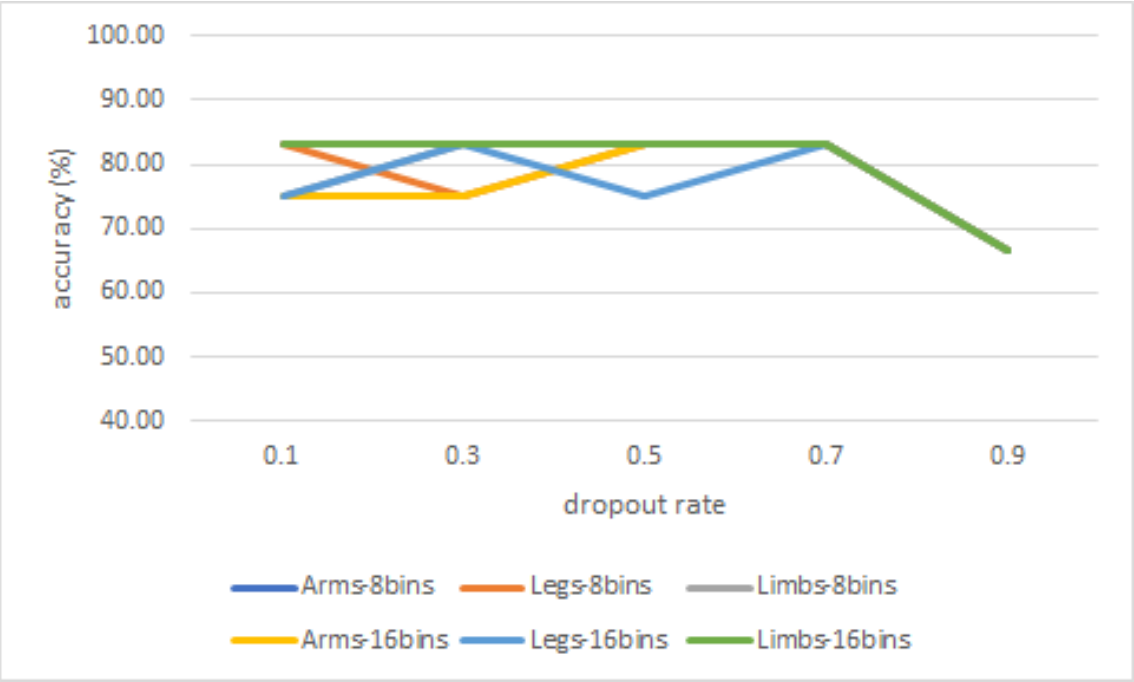


Figure 6.8: *FCNet* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

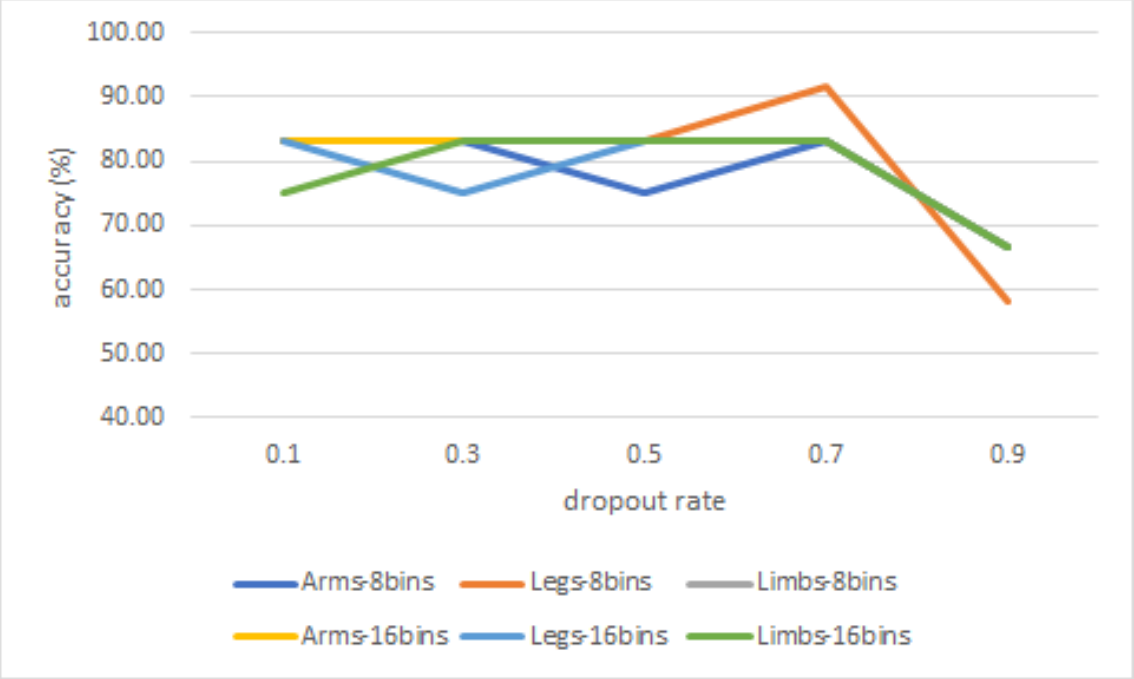


Figure 6.9: *Conv1D-1* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

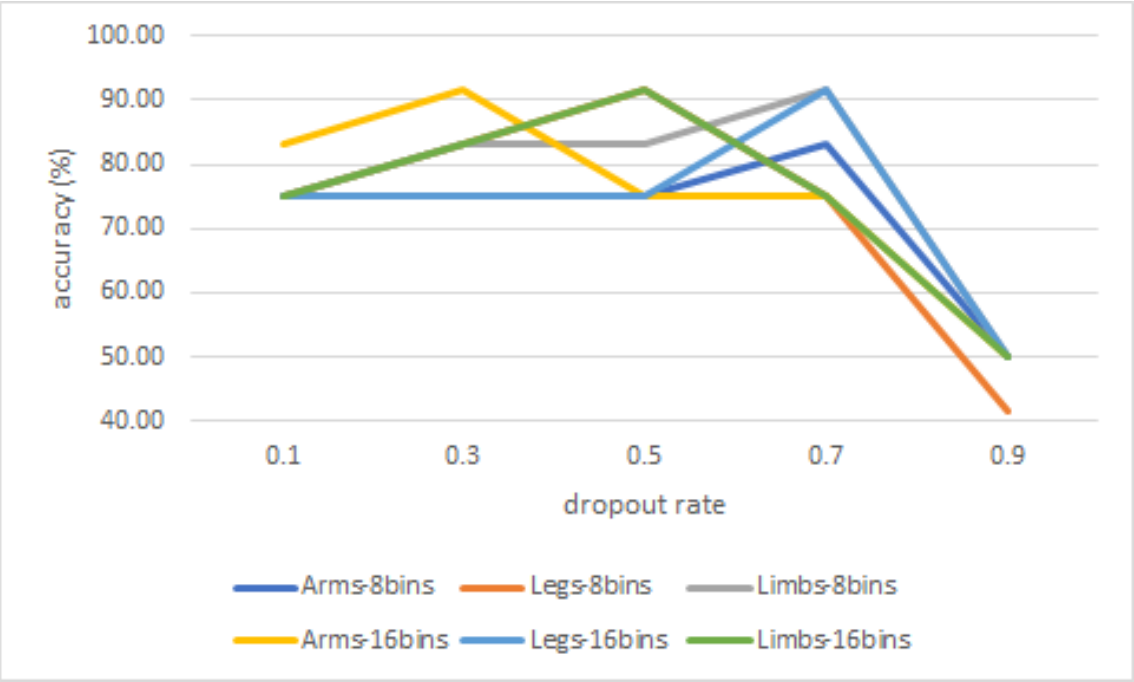


Figure 6.10: *Conv1D-2* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

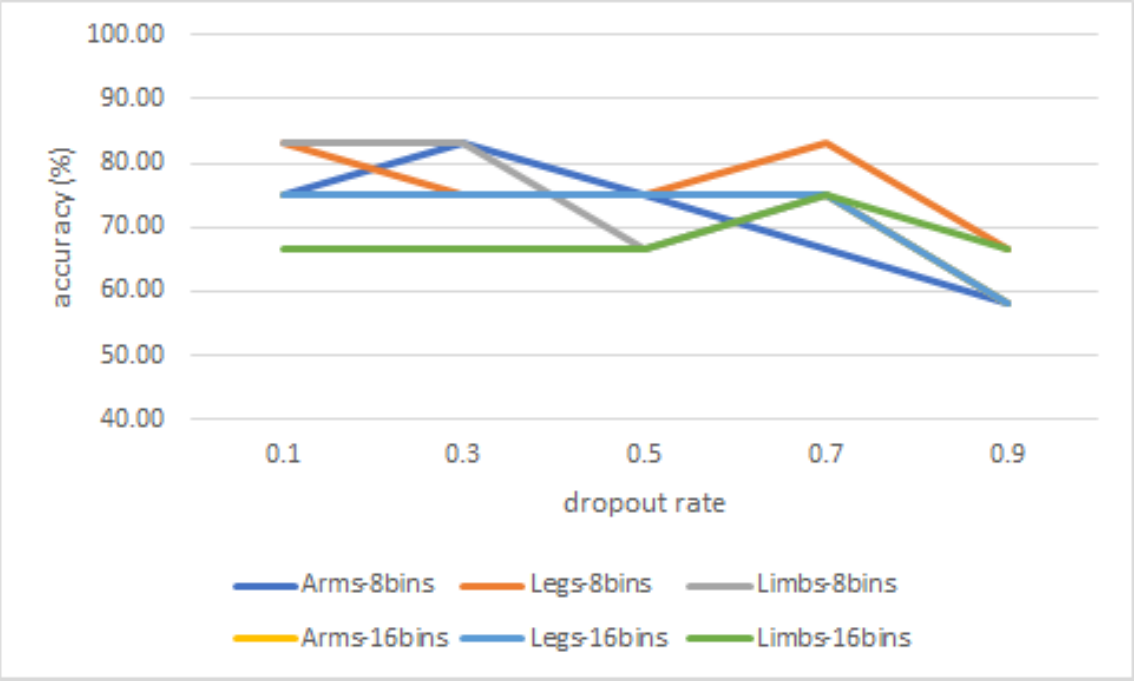


Figure 6.11: *Conv2D-1* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

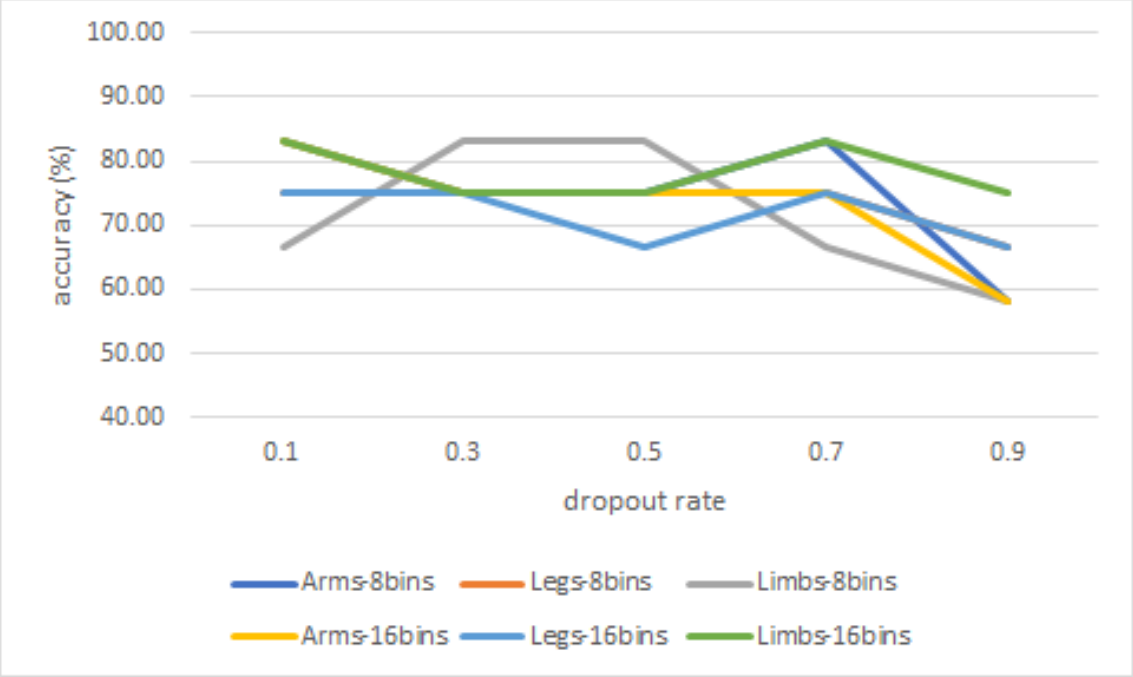


Figure 6.12: *Conv2D-2* ablation testing using the fused HOJO2D and HOJD2D feature sets: The effect of dropout rate on classification performance.

6.2.4 Improved Machine Learning Framework Classification Performance

In our next series of experiments, we evaluate the proposed improved traditional machine learning framework discussed in Section 6.1.3. In addition to evaluating the classification performance of each of the individual features we also evaluate the performance of the fused features for comparison as discussed in Section 5.2. In this case, the fusion process concatenates features into two separate feature sets: *pose-based* features, and *velocity-based* features, which are used for classification. Additionally, we then fuse the *pose-based* and *velocity-based* feature sets together to create a more robust representation, since we have found that fused features have typically demonstrated the best performance. We then evaluate this against both the individual features, and the selected baselines from the literature as discussed in Section 5.1; on both the publicly available MINI-RGBD dataset [150], and the RVI-38 dataset, a challenging new video dataset gathered as part of routine clinical care. We again employ leave-one-out cross-validation throughout and further evaluate the performance of each method by using several additional metrics, as discussed in Section 6.2.1. We provide a full discussion on our observations and interpretations of the results, as well as a full analysis of the proposed features and classification algorithm hyper-parameter optimisation.

Table 6.7: Classification accuracy comparison between our proposed supplementary features and the selected baselines on the MINI-RGBD dataset [150].

Feature	**Class.**	**AC**	**SE**	**SP**	**F1**	**MCC**
CX_{m} [115]	Ens	83.33	50.00	**100.00**	66.67	63.25
CX_{SD} [115]	Ens	83.33	75.00	87.50	75.00	62.50
CY_{m} [115]	LDA	33.33	**100.00**	0.00	50.00	0.00
CY_{SD} [115]	k=3	75.00	75.00	75.00	66.67	47.81
$\mathrm{Q_m}$ [115]	Ens	58.33	50.00	62.50	44.44	11.95
$\mathrm{Q_{SD}}$ [115]	k=3	75.00	75.00	75.00	66.67	47.81
CPP [128]	Tree	66.67	75.00	62.50	60.00	35.36
AMD [116]	LDA	91.67	**100.00**	87.50	88.89	83.67
MWC [116]	LDA	83.33	75.00	87.50	75.00	62.50
RF [116]	LDA	91.67	**100.00**	87.50	88.89	83.67
$\mathrm{FFTx_m}$ [110]	Tree	83.33	75.00	87.50	75.00	62.50
$\mathrm{FFTx_{SD}}$ [110]	Ens	58.33	50.00	62.50	44.44	11.95
$\mathrm{FFTy_m}$ [110]	Tree	91.67	75.00	**100.00**	85.71	81.65
$\mathrm{FFTy_{SD}}$ [110]	Tree	75.00	75.00	75.00	66.67	47.81
$\mathrm{FFT_m}$ [110]	Tree	83.33	75.00	87.50	75.00	62.50
$\mathrm{FFT_{SD}}$ [110]	LR	75.00	75.00	62.50	72.73	59.76
MCI [153]	n/a	91.67	**100.00**	87.50	88.89	83.67
CA [154]	DNN	91.67	-	-	-	-
BPB [156]	DNN	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
STAM [155]	DNN	91.67	**100.00**	87.50	88.89	83.67
HOJO2D	Ens	91.67	75.00	**100.00**	85.71	81.65
HOJD2D	Ens	83.33	75.00	87.50	75.00	62.50
FFT-JO	Ens	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
FFT-JD	LR	91.67	75.00	**100.00**	85.71	81.65
HOAD2D	LR	66.67	**100.00**	50.00	66.67	50.00
HORJO2D	LDA	91.67	75.00	**100.00**	85.71	81.65
HORJAD2D	Ens	83.33	75.00	87.50	75.00	62.50
Pose	Ens	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Velocity	Ens	91.67	**100.00**	87.50	88.89	84.32
Pose & Vel.	Ens	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Classification Performance on Individual Features

From Table 6.7, we observe that our proposed feature FFT-JO achieved a 100.00% classification accuracy on the MINI-RGBD dataset. Only one of the 20 baseline methods (BPB [156]) evaluated achieved this perfect classification result in our tests, highlighting the remarkable performance of this new feature. Encouraging results are also obtained using our other frequency-based feature FFT-JD, with 91.67% classification accuracy, 85.71% F1 score, and 81.65% MCC. This performance is higher than all of the 20 baselines in the experiments, with the exception of AMD (F1:88.89%, MCC:83.67%), RF (F1:88.89%, MCC:83.67%), FFT-Y_m (F1:85.71%, MCC:81.65%), MCI (F1:88.89%, MCC:83.67%), and STAM (F1:88.89%, MCC:83.67%). Similarly, our newly proposed HORJO2D feature achieved the same performance of 91.67% classification accuracy, 85.71% F1 score, and 81.65% MCC. However, the other relative orientation-based feature HORJAD2D is not performing as well on this dataset with 83.33% classification accuracy, an F1 Score of 75% and MCC of 62.50%. Although this performance still outperforms most of the baselines, the noticeably lower specificity (87.50%) results in a lower overall classification performance for this feature. For the angular displacement based feature HOAD2D, an average performance is obtained on this dataset with an F1 score of 66.67%, matching or outperforming 8 of 20 baselines.

For the RVI-38 dataset, we note a general drop in performance due to the challenging nature of the dataset, as shown in Table 6.8. This is particularly noticeable in the baseline methods where we see a significant drop for each baseline, with the exception of MWC (F1:75%, MCC:62.50%). This drop is most likely associated with the challenging nature of the captured data and the full frame analysis of these methods. We are seeing that methods which are able to deal with external influences better, such as the pose-based methods, are generally producing more accurate results. This is also reflected in the results produced using our proposed individual features. In this setting we note that the HOAD2D feature is performing particularly well, representing the strongest individual feature on this dataset, recording the highest F1 Score (83.33%) and MCC (80.21), along with the joint highest accuracy (94.74%) and sensitivity (83.33%). The reworked HOJO2D (F1:72.73%, MCC:68.54) and HOJD2D (F1:80.00%, MCC:79.21%) again perform well, showing the robustness of these improved features. The HORJO2D feature is also performing well, with an accuracy of 92.11%, F1 score of 76.92%, and MCC of 72.51%. We note that FFT-JD is once again

Table 6.8: Classification accuracy comparison between our proposed supplementary features and the selected baselines on the RVI-38 dataset.

Feature	Class.	AC	SE	SP	F1	MCC
CX_{m} [115]	LR	50.00	**83.33**	43.75	34.48	20.20
CX_{SD} [115]	Ens	68.42	33.33	75.00	25.00	6.90
CY_{m} [115]	k=3	84.21	50.00	90.63	50.00	40.63
CY_{SD} [115]	LR	63.16	66.67	65.63	36.36	21.54
Q_{m} [115]	LR	52.63	50.00	53.13	25.00	2.28
Q_{SD} [115]	k=1	86.84	50.00	93.75	54.44	47.19
CPP [128]	Ens	84.21	50.00	90.63	50.00	40.63
AMD [116]	LDA	83.33	50.00	**100.00**	66.67	63.25
MWC [116]	Tree	83.33	75.00	87.50	75.00	62.50
RF [116]	LDA	84.21	66.67	87.50	57.14	48.45
FFT-X_{m} [110]	Ens	84.21	50.00	90.63	50.00	40.63
FFT-X_{SD} [110]	LR	63.16	66.67	62.50	36.36	21.54
FFT-Y_{m} [110]	k=1	81.58	33.33	90.63	36.36	25.84
FFT-Y_{SD} [110]	LDA	55.26	50.00	56.25	26.09	4.58
FFT_{m} [110]	Tree	84.21	50.00	90.63	50.00	40.63
FFT_{SD} [110]	LR	42.11	66.67	37.50	26.67	3.15
BPB [156]	DNN	84.21	33.33	93.75	40.00	32.18
STAM [155]	DNN	81.58	33.33	90.63	36.36	25.85
HOJO2D	Ens	92.11	66.67	96.88	72.73	68.54
HOJD2D	Ens	94.74	66.67	**100.00**	80.00	79.21
FFT-JO	LDA	84.21	**83.33**	84.38	62.50	56.07
FFT-JD	Ens	92.11	66.67	96.88	72.73	68.54
HOAD2D	Ens	94.74	**83.33**	96.88	83.33	80.21
HORJO2D	Tree	92.11	**83.33**	93.75	76.92	72.51
HORJAD2D	LR	86.84	66.67	90.63	61.54	53.89
Pose	Ens	94.74	**83.33**	96.88	83.33	80.21
Velocity	Ens	94.74	66.67	**100.00**	80.00	79.21
Pose & Vel.	Ens	**97.37**	**83.33**	**100.00**	**90.91**	**89.89**

performing well, with an accuracy of 92.11%, an F1 score of 72.73%, and MCC of 68.54%. We also observe that whilst FFT-JO and HORJA2D achieve a reasonable performance on the RVI-38 dataset, with 84.21% and 86.84% classification accuracy respectively, the F1 and MCC scores are lower than our other proposed features. However, whilst the scores for these features are not class leading, they are still higher than those achieved by 16 of the 18 baseline methods evaluated in this setting.

Classification Performance on Feature Fusion

From Table 6.7, we observe that on the MINI-RGBD dataset the pose-based fusion is extracting the strongest feature representation and retaining the perfect classification performance provided

by the FFT-JO individual feature. Our evaluation also suggests that whilst the pose-based fused features are generally outperforming the velocity-based features, fusing both of these feature sets further improves performance on both the MINI-RGBD dataset (F1: 100%, MCC 100.00%) and the RVI-38 dataset (F1: 90.91%, MCC: 89.89%). From Table 6.8, we note that on the RVI-38 dataset, the strengths from each feature set combine to provide this improved overall classification performance, with the higher sensitivity found in the pose-based features (83.33%) and the higher specificity found in the velocity-based features (100.00%) directly translating to the concatenated fusion of these feature sets. This observation aligns well with our feature design, which looks to incorporate the combined positional, directional, postural, and transitory information specified in the GMA guidelines. We also note that on the fused features we are seeing a consistently high performance using the ENS classifier, with the best results obtained on both datasets for all fused feature sets using this classification method. The evaluation metrics also highlight the robustness of the proposed feature fusion method, given that only one sample video was misclassified across both datasets.

Hyperparameter Optimisation

To refine the framework performance, we further investigate hyper-parameter optimisation using Bayesian Optimisation [162]. By using Bayesian Optimisation we efficiently utilise past results to inform the decisions made relating to minimising the cross-validation loss. The optimisation searches over the ensemble aggregation methods for binary classification to produce an output with the minimum estimated cross-validation loss, and due to its ability to learn from previous iterations, is seen as more effective than other optimisation methods such as grid search and random search. Using this method, we evaluate the results of optimisation in an informed manner, by tuning the learning rate, the number of learning cycles, the minimum observations per leaf, and the maximum number of branch nodes, to minimize the cross-validation loss of the classifier. We conducted the hyperparameter optimisation on the ENS classifier using the fused features for both pose and velocity, and carried optimisation out using both the RVI-38 and MINI-RGBD datasets.

We present several plot representations of our Bayesian Optimisation in Figure 6.13. In the plots, the number of function evaluations relates to the iteration number of the objective function, the

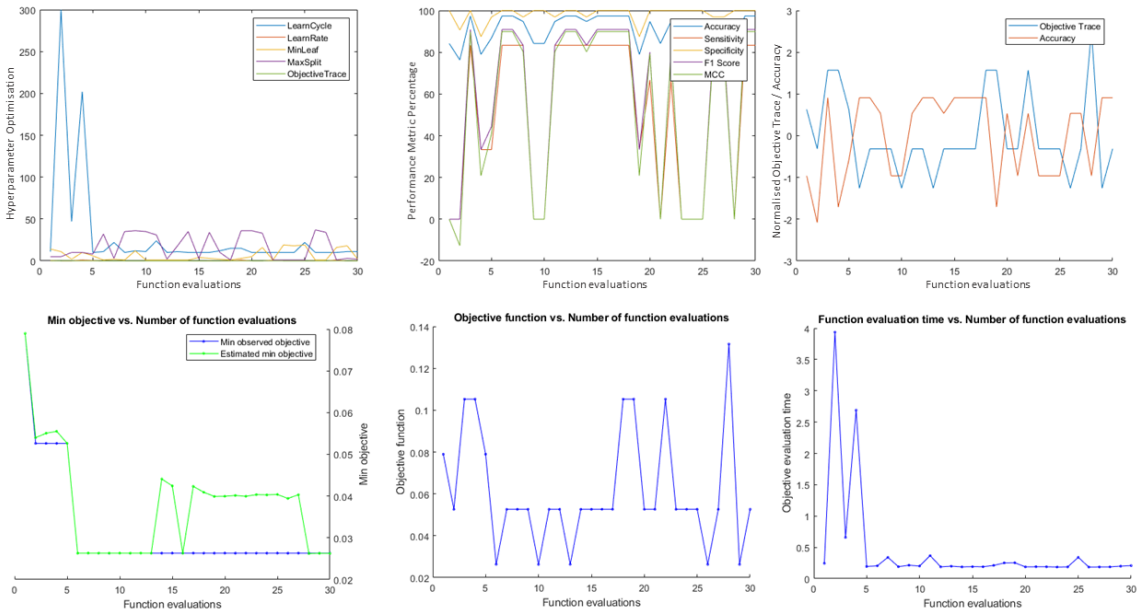


Figure 6.13: Hyperparameter optimisation plots on the RVI-38 dataset

min objective is the minimum value that the objective function has reached up to the current iteration, and the estimated minimum objectives are the mean values of the posterior distribution of the Gaussian process model of the objective function [163]. We also map the hyper-parameter variables to provide visual interpretation of the classification performance metrics relative to the optimal hyper-parameters. In the plots shown in we generally observe improved performance in line with a reduction in the calculated objective function values.

In Figure 6.14, we show the results of each iteration along with the associated hyperparameter values, followed by the finalised model hyperparameters that produced the best classification performance. In this setting, we found the optimal hyper-parameters to be: 0.1045800 learning rate, 11 learning cycles, 1 minimum observation per leaf, and 32 split branch nodes, providing an objective function of 0.026316 and an accuracy of 100% on the MINI-RGBD dataset, and 97.37% on the RVI-38 dataset, per our reported results.

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	NumLearningCycles	LearnRate	MinLeafSize	MaxNumSplits
1	Best	0.078947	0.24345	0.078947	0.078947	11	0.45045	14	5
2	Best	0.052632	3.9358	0.052632	0.053998	300	0.0035742	11	5
3	Accept	0.10526	0.65968	0.052632	0.055054	47	0.05441	2	10
4	Accept	0.10526	2.6906	0.052632	0.055525	202	0.97296	10	10
5	Accept	0.078947	0.19377	0.052632	0.052635	10	0.40361	6	8
6	Best	0.026316	0.20383	0.026316	0.02632	11	0.1045800	1	32
7	Accept	0.052632	0.34022	0.026316	0.026324	22	0.0010157	2	3
8	Accept	0.052632	0.19107	0.026316	0.026325	10	0.0022845	1	35
9	Accept	0.052632	0.21379	0.026316	0.026326	12	0.0012333	12	36
10	Accept	0.026316	0.20077	0.026316	0.026317	11	0.0013515	1	35
11	Accept	0.052632	0.36793	0.026316	0.026314	24	0.0012257	1	31
12	Accept	0.052632	0.18761	0.026316	0.026315	10	0.0012061	1	2
13	Accept	0.026316	0.19812	0.026316	0.026311	11	0.0012245	1	18
14	Accept	0.052632	0.18729	0.026316	0.044088	10	0.0011662	1	35
15	Accept	0.052632	0.19101	0.026316	0.042452	10	0.013145	4	3
16	Accept	0.052632	0.18957	0.026316	0.026316	10	0.068331	3	34
17	Accept	0.052632	0.21082	0.026316	0.04232	12	0.0011046	2	10
18	Accept	0.10526	0.25026	0.026316	0.04091	15	0.035457	1	1
19	Accept	0.10526	0.25245	0.026316	0.039918	15	0.98303	3	36
20	Accept	0.052632	0.18795	0.026316	0.039951	10	0.001014	5	36

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	NumLearningCycles	LearnRate	MinLeafSize	MaxNumSplits
21	Accept	0.052632	0.18897	0.026316	0.04015	10	0.024247	16	33
22	Accept	0.10526	0.18894	0.026316	0.039941	10	0.73279	1	1
23	Accept	0.052632	0.18539	0.026316	0.040343	10	0.0011703	19	1
24	Accept	0.052632	0.18633	0.026316	0.040289	10	0.034421	18	1
25	Accept	0.052632	0.34036	0.026316	0.040377	22	0.010948	19	1
26	Accept	0.026316	0.18469	0.026316	0.039428	10	0.16316	1	37
27	Accept	0.052632	0.18619	0.026316	0.040311	10	0.019934	1	34
28	Accept	0.13158	0.18699	0.026316	0.026232	10	0.9685	16	1
29	Accept	0.026316	0.19872	0.026316	0.026284	11	0.0012162	18	3
30	Accept	0.052632	0.20856	0.026316	0.026284	11	0.13603	2	2

```
Optimization completed.
MaxObjectiveEvaluations of 30 reached.
Total function evaluations: 30
Total elapsed time: 31.0726 seconds
Total objective function evaluation time: 13.1511

Best observed feasible point:
    NumLearningCycles    LearnRate    MinLeafSize    MaxNumSplits
    _________________    _________    ___________    ____________

           11             0.0010458         1              32

Observed objective function value = 0.026316
Estimated objective function value = 0.026348
Function evaluation time = 0.20383

Best estimated feasible point (according to models):
    NumLearningCycles    LearnRate    MinLeafSize    MaxNumSplits
    _________________    _________    ___________    ____________

           11             0.0012245         1              18

Estimated objective function value = 0.026284
Estimated function evaluation time = 0.20096

Best: The average Ensemble classification accuracy is 97.3684%.
Best: The Ensemble Sensitivity is 83.3333%.
Best: The Ensemble Specificity is 100%.
Best: The Ensemble F1 Score is 90.9091%.
Best: The Ensemble MC Score is 89.8933%.
```

Figure 6.14: Hyperparameter optimisation results on the RVI-38 dataset

6.2.5 Analysis of the Proposed Features

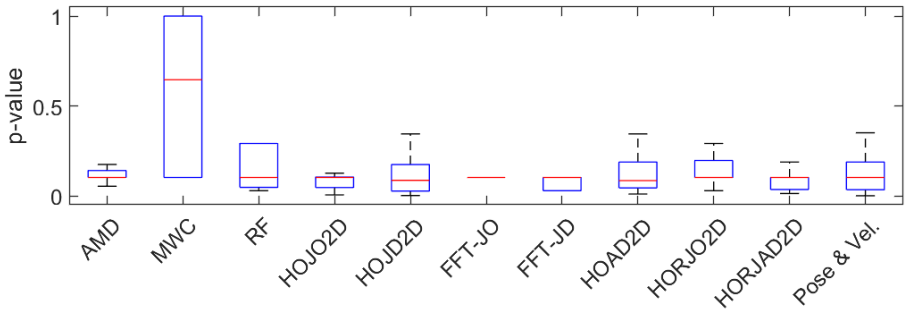
Chi-squared Tests

To further evaluate the discriminative power of the newly proposed features, chi-square tests are used for testing if the predictor variables (i.e. the multi-dimensional features proposed in this work) and the response variable (i.e. the label of each video) are related. In particular, such tests have been widely used for feature selection and are thus able to reflect the quality of the features we propose. We conducted the chi-square tests on both the MINI-RGBD and the RVI-38 datasets to highlight the differences between these two datasets. Specifically, the p-value for each predictor variable is calculated and the median values are reported in Table 6.9. Here, we consider the predictor variables as *significant predictor (sp)* if $p < 0.05$. Since the features used in the experiments are mostly multi-dimensional, the dimensionality (dim.), number of *sp* (# *sp*) and percentage of *sp* (% *sp*) are also reported in Table 6.9. We also include the top performing baselines from our experiments (i.e. AMD, MWC and RF proposed in [116]) in this analysis to further highlight the effectiveness of the proposed features.

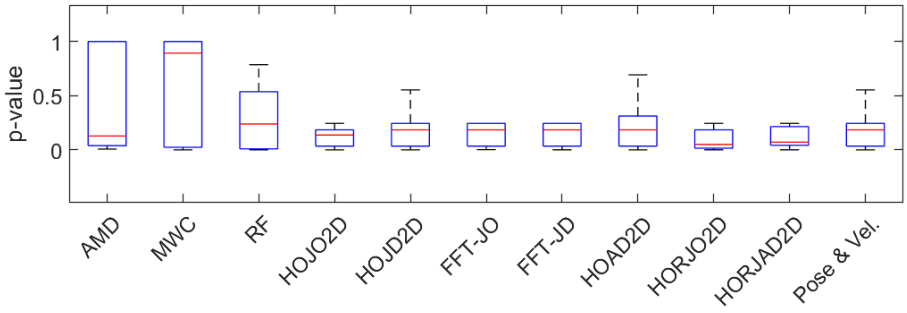
Table 6.9: The p-values of the features computed from chi-square tests on the MINI-RGBD and RVI-38 datasets.

		MINI-RGBD			RVI-38		
Feature	**Dim.**	**Median**	**#** *sp*	*% sp*	**Median**	**#** *sp*	*% sp*
AMD [116]	8	0.1020	0	0%	0.1280	2	25%
MWC [116]	18	0.6456	0	0%	0.8931	6	33.33%
RF [116]	16	0.1020	4	25.00%	0.2388	6	37.50%
HOJO2D	64	0.1020	18	28.13%	0.1370	10	31.25%
HOJD2D	128	0.0860	52	**40.63%**	0.1849	45	35.16%
FFT-JO	48	0.1020	8	16.67%	0.1849	18	32.14%
FFT-JD	64	0.1020	23	35.94%	0.1849	18	28.13%
HOAD2D	64	**0.0847**	17	26.56%	0.1849	22	34.38%
HORJO2D	32	0.1020	5	15.63%	**0.050**	14	**43.75%**
HORJAD2D	32	0.1020	9	28.13%	0.071	8	25%
Pose & Vel.	792	0.1020	228	28.78%	0.1849	260	32.83%

From Table 6.9, it can be seen that HOJD2D and HORJO2D achieve the highest % *sp* on the MINI-RGBD and RVI-38 datasets, respectively. On the MINI-RGBD dataset, our proposed features are producing the same or lower median p-values when compared with AMD, MWC and RF. On the RVI-38 dataset, HORJO2D and HORJAD2D achieved significantly lower median p-values than the other top performers. Whilst AMD performed better than some of our proposed features in



(a) MINI-RGBD



(b) RVI-38

Figure 6.15: Boxplots of the p-values of different features on each dataset.

terms of the median p-values, we observe that the results from the 2 datasets demonstrate the robustness of our proposed features, since our features achieve more consistent results.

To better visualize the distribution of the p-values, boxplots of the p-values of different features on both the MINI-RGBD and RVI-38 datasets are shown in Figure 6.15a and 6.15b, respectively. In particular, the maximum, minimum, first quartile, third quartile and median (red line) values are illustrated in the figures. It can be seen that the majority (from the first to third quartile) of the predictor variables in our proposed features are producing a small range with low p-values. This indicates the majority of our proposed features are of higher importance and quality when compared with the the top performing baseline methods.

Histogram Variance Tests

As an additional means of assessing the discriminative power of the proposed features, we also analyze the variance of each histogram feature extracted from different body parts on both of the datasets. The variance is evaluated since the scalar value indicates whether the values are evenly spread across the range or only have a small number of bins with high values, aligning with the

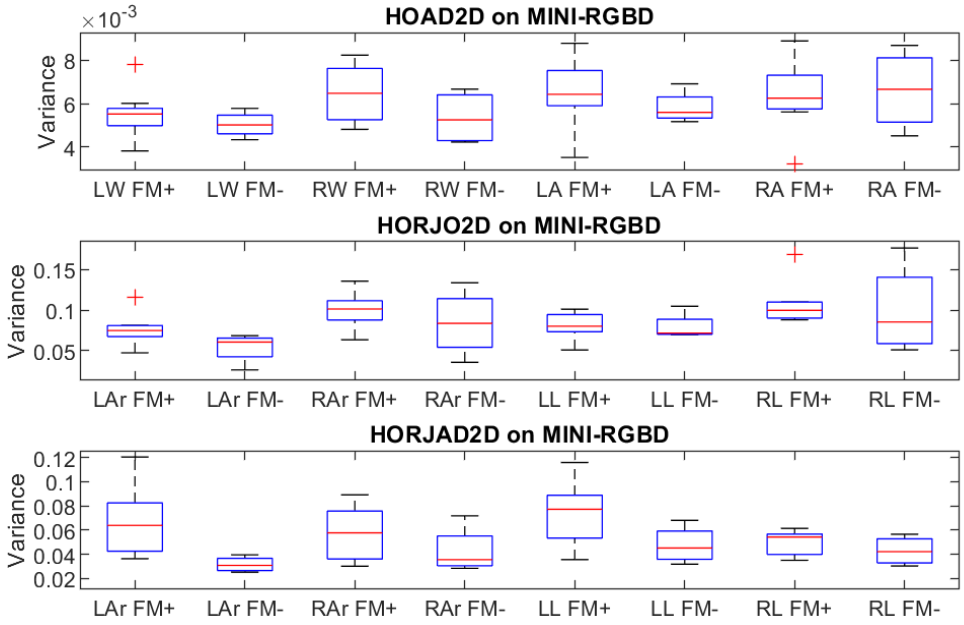


Figure 6.16: Boxplots of the variance computed from the proposed histogram feature. The results are grouped by FM+/FM- samples extracted from different body parts on the MINI-RGBD dataset.

design of our features, as explained in Section 5.2. In this experiment, we separate the videos into FM+ and FM- groups for each dataset, and present the variance computed from each of the histogram features using box plots (refer to Figure 6.16 for MINI-RGBD and Figure 6.17 for RVI-38), which include the maximum, minimum, first quartile, third quartile, median values (the red line) and outliers (the '+' signs).

For HOAD2D, we selected 4 key body parts, *LW: Left Wrist, RW: Right Wrist, LA: Left Ankle and RA: Right Ankle*, as they represent the limb movements. In general, the variance of HOAD2D is consistent across the 2 datasets in which the medians of the FM+ samples are higher than the FM-. This suggests that the FM+ samples have large values in a small number of bins, which refers to a narrower range of angular displacements (speed). This shows the FM+ samples are generally having smoother movement with less speed change over the FM- samples. In Figure 6.18a, we selected a sample HOAD2D feature with median variance to illustrate the difference in the distribution of the angular displacement between FM+ and FM- samples.

For the relative joint orientation based features, HORJO2D and HORJAD2D, we follow the design by computing the variance for the histograms extracted from the 4 limbs, *LAr: Left Arm, RAr: Right Arm, LL: Left Leg and RL: Right Leg*. A similar trend can generally be observed for both of the features, as the FM+ samples have a higher median value than the FM- samples on

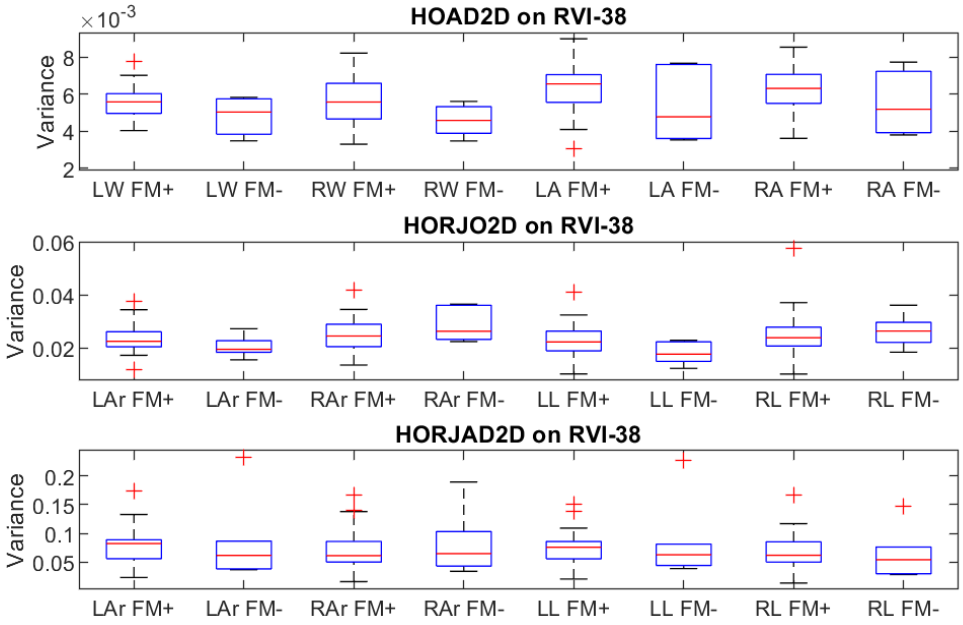
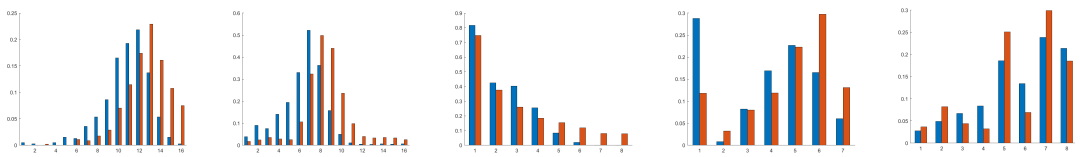


Figure 6.17: Boxplots of the variance computed from the proposed histogram feature. The results are grouped by FM+/FM- samples extracted from different body parts on the RVI-38 dataset.

the MINI-RGBD dataset, although the difference between the FM+ and FM- sample is less pronounced on the RVI-38 dataset. This highlights the performance of the relative joint orientation based features is better on the MINI-RGBD dataset when compared with the RVI-38 dataset. A sample of the HORJAD2D feature with the median variance is illustrated in Figure 6.18c. It can be observed that the relative joint orientation displacements tend to have smaller values for FM+ when compared with FM-. This shows the FM+ samples have less changes in angular displacement resulting in smoother body movements. In Figure 6.18b, the sample HORJO2D feature with the median variance shows different distributions in FM+ and FM- samples.



(a) HOAD2D (b) HORJO2D (c) HORJAD2D (d) FFT-JO (e) FFT-JD

Figure 6.18: Visualizing examples of the proposed histogram features. FM+ and FM- samples are colored in blue and orange, respectively.

For FFT-JO and FFT-JD, the variance of the histogram is less informative as the distribution of the values towards the lower or higher frequency bands is more crucial in representing the underlying

motion pattern. Here, median value histograms of FFT-JO and FFT-JD are illustrated in Figure 6.18d and 6.18e, respectively. The FFT-JO of the FM+ sample shows more low-frequency angular movement at the joints, which aligns with the anticipated smoother motion characteristics. Conversely, the FM- sample shows an opposite pattern with more high-frequency movements. The FFT-JD histograms show a different trend when compared with FFT-JO and the FFT features [110] extracted from the x-y trajectory of each limb, by having higher values towards the high-frequency bins for both FM+ and FM-. This is because the FFT-JD is primarily capturing the pattern of the change of joint displacement over time. Nevertheless, we observe a different distribution of the values in the histograms for FM+ and FM- samples.

6.3 Discussion

In our an initial experiments we examined the validity of using pose-based methods to analyse and classify infant part-based body movements from video footage. To do this we evaluated the effectiveness of our proposed pose-based features, HOJO2D and HOJD2D, by conducting a series of experiments to classify the body-part movement into two categories, 'normal' and 'abnormal'. Encouraging results were obtained, with high accuracy (91.67%) achieved in several different settings with the ENS classifier, with some part-based settings attaining 100% accuracy.

The proposed features demonstrate the benefits of pose-based analysis by dealing with information which can traditionally affect classification results, such as loose clothing, illumination changes, body-part variation, and background clutter. Our initial results suggest that our histogram-based method is capable of mapping infant movements associated with the GMA for further classification. This study also highlights that this method of motion analysis is viable on an infant population and as such represents a step towards the automated classification of infant movements for the prediction of later CP using the GMA.

In our next series of experiments we evaluated our proposed deep learning based frameworks for this same classification task. The proposed frameworks were evaluated and compared with our previous methods. Our experimental results show that the proposed fully connected neural network *FCNet* performed robustly across different feature sets. Furthermore, the proposed 1D convolutional neural network architectures demonstrated excellent performance in handling features

in higher dimensionality, with more consistent performance throughout. However, the implementation of deep learning frameworks still affects the interpretability of the model, and whilst we use these purely for classification, they remain less human explainable than traditional machine learning models. Also, given that the quantity of video sequences used in these experiments is relatively small, and since deep learning methods are still typically reliant on having more data than traditional machine learning approaches, there is still merit in pursuing traditional methods as a viable alternative.

Given the importance of clinical interpretability in our pipeline, for our next experiments, we evaluated the integration of several new interpretable pose-based features, as well as several additional traditional machine learning classifiers. Our proposed features were developed relating directly to the criteria associated with the GMA checklist [47] and HINE Optimality Score [164], as discussed in Section 5.2. Additionally, based upon our previous improved performance through feature fusion, we fused these features together to produce a more robust representation of infant body movement for classification. We compared these features with several other methods from the literature by re-implementing them for assessment using shared datasets.

From our evaluation, we observe that our individual features are typically performing at a similar level to or exceeding that of the best methods proposed in the related works. However, we note that the fusion of our proposed features is providing state-of-the-art performance across both of the datasets used in our experiments. Our results on the RVI-38 dataset in particular represent a particularly robust performance given the difficulty of the associated dataset, with only one misclassified video. In this case the misclassified video was one of the positive samples which, in practice presents a greater issue than a misclassified negative sample, but this performance on a particularly challenging dataset is extremely encouraging. Additionally, our pose and velocity-based method is simpler to understand, retains understandable information, and has less parameters to tune than the related methods, making it more accessible in a clinical setting. We also suggest that, due to the relative assessment of joint motion, our framework is better able to deal with camera movement, changes in resolution, variable infant sizes, and larger motion changes between frames. As such, we suggest that our proposed pose-estimation based approaches provide several advantages over the previously proposed methods in data acquisition and analysis, whilst simultaneously providing state-of-the-art performance.

6.4 Concluding Remarks

In this chapter, we have discussed how we evaluated the performance of our proposed features and classification frameworks, and the metrics used in our experiments. We examined how our proposed methods compared with baselines derived from the literature and discussed the advantages of our frameworks over the existing methods. We found that our proposed features and frameworks achieved state-of-the-art performance across two datasets, whilst retaining significantly greater clinical interpretability than the related works. We also suggested that since the video sequences used in our earliest works are synthetic, the appearance of the images used as an input for the OpenPose framework may differ slightly from that of real-world video data. As such, we also evaluated our improved framework against methods from the literature using data captured from real-world video footage of infants.

We have established the viability of our pose-based method for infant motion analysis. However, through our collaborative relationship with healthcare professionals, it has become also apparent that, in addition to high accuracy, explainability and interpretability at all stages in the automated predictive framework is key. As such, in the next chapter we investigate the feasibility of including visualisation components to inform clinicians of pertinent information throughout the classification process to further enhance interpretability.

Chapter 7

Visualisation to Enable Explainable AI

In this chapter we consider the importance of explainable AI for the automated prediction of CP. We discuss our proposed approach to enable explainable AI, and provide details of our proposed framework relating to the generation of spatio-temporal features for classification and visualisation to improve clinical interpretablity. We then evaluate the effectiveness of our proposed framework both quantitatively and qualitatively, and discuss how our method fits withing the clinical setting.

7.1 The Importance of Explainable AI for Clinical Applications

The studies discussed thus far suggest that an automated system could potentially help to reduce the time and cost associated with current manual clinical assessments, and also assist clinicians in making earlier and more confident diagnoses by providing additional information about the assessed infant movements. Whilst machine learning-based frameworks have obtained excellent performance in a wide range of visual understanding tasks, these methods are also not without their setbacks. One of the main issues with using machine-learning approaches in the medical domain is the problem of interpretable AI. Models are often seen as 'black boxes' in which the underlying structures can be difficult to understand, since most of the classification frameworks only output the predicted label without specifying exactly what influences the classification decision. Whilst this is acceptable in typical computer vision tasks, it is less preferable in healthcare applications, since it is essential for the clinicians to verify the prediction as well. There is therefore an increasing requirement for the mechanisms behind why systems are making decisions to be transparent, understandable and explainable.

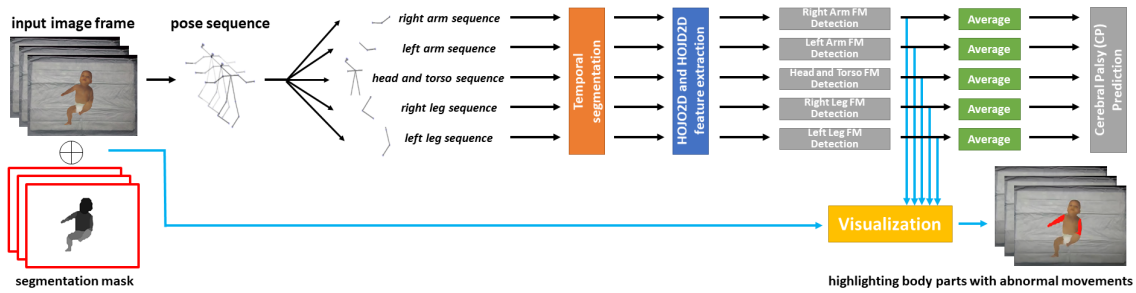


Figure 7.1: The overview of the proposed prediction and visualization framework.

7.2 The Proposed Approach

In this work, our framework again takes the processed video data as the input, but this time analyzes the movement of individual body parts from temporal segments to determine if FMs are present (FM+) or absent (FM-), subsequently identifying normal or abnormal general movements from sub-sections of the sequence. To make our proposed framework fully interpretable, an important aspect is the integration of an automatically generated visualization capable of relaying pertinent information to the assessor. The visualization highlights body-parts which are showing

movement abnormalities, and are subsequently providing the most significant contribution towards the classification result. This new approach not only allows us to take advantage of additional body-part specific information as a means of improving classification accuracy, but also affords us the opportunity to include finer detail than previous methods allowing for greater interpretability and enhanced classification performance on the proposed HOJO2D and HOJD2D features.

Pose-based motion features

The first step of our proposed framework is to make use of the extracted pose-based HOJO2D and HOJD2D features, which we established the validity of in Chapter 6. In this work, an early fusion (i.e. concatenation) of HOJO2D and HOJD2D is used as the input, since we have found that fused features have consistently demonstrated the best performance.

Spatiotemporal Fidgety Movement Detection

In order to detect the presence of FMs spatiotemporally, the motion features have to be extracted from 1) different body-parts and 2) different temporal segments individually. As such, we propose motion feature extraction from 5 different body-parts in the spatial domain, namely *left arm, right arm, left leg, right leg, and head-torso*. For the temporal domain, we compute HOJO2D and HOJD2D features (8 bins) for the 5 body parts from every 100-frame segment. In doing so, each video is represented by multiple histogram-based motion features accordingly. For example, a 1000-frame video will be represented by 50 fused features of HOJO2D and HOJD2D.

In this work, we formulate the FMs detection problem as a binary classification. Since each video is annotated with FM+ or FM-, we label all the fused features extracted according to the holistic annotation of the video. When training the classifier all features are used, while the temporal location information is not used. In other words, no matter whether the features are extracted from the beginning or near the end of the video, they will be used to train a single classifier. This proposed approach provides distinct advantages over previous methods, i.e. 1) the classifier will be trained by more data samples rather than using only one histogram representation for the whole video, and 2) a focus on the presence/absence of FMs while ignoring the temporal information when training the classifier.

We make use of the ENS classifier, which consists of a range of classifiers to boost the performance of the classification results. Given the multiple fused features extracted from a video, all the features will be classified as FM+ or FM-, this information is then used in visualizing the results (Section 7.1). As the features were extracted in sequential order in the temporal domain, the classification result on each histogram-based motion feature is essentially detecting FMs spatiotemporally.

Late Fusion for Cerebral Palsy Prediction

While the method presented in Section 7.2 provides precise information on the presence/absence of FMs spatiotemporally, directly using all motion features as a CP prediction for the whole video will result in sub-optimal performance since the temporal ordering is less important in the GMA than the presence/absence of sustained FMs at any point in the sequence. To tackle this problem, we propose representing each of the 5 body parts using a single scalar score s, with this being the average score of the classification result (FM+ as 0 and FM- as 1) across all temporal segments for each body-part. Therefore, the range of s will be between 0 and 1.

Here, we propose to use a late fusion approach to train an ensemble classifier for cerebral palsy prediction. Specifically, each video is represented by using the 5 scores obtained from the body parts. The binary classifier will predict whether the motion in the video is considered *normal* or *abnormal*.

Visualisation

In order to make our proposed framework more interpretable, we include a visualization module that highlights the body-parts that are contributing to the classification decision. Our proposed method highlights the body-parts in *red* to indicate the *absence of fidgety movements* based on the scores computed in the body part abnormality detection explained in Section 7.2, providing clinicians with an intuitive visualization such as the examples illustrated in Figure 7.3.

To extract body part information from an input image, the CDCL [25] pre-trained body segmentation model is used in this work. The body is segmented into 6 parts; head, torso, upper arms, lower arms, pelvis and upper legs, and lower legs. An example of the segmentation result is illustrated in Figure 7.2. Specifically, given an input infant image, CDCL [25] returns an image mask for seg-

Figure 7.2: Segmentation result obtained using CDCL [25].

mentation. To align with the 5 body parts used in this work, we separate the segmentation masks for the arms and legs into the left and right masks. Here, k-means clustering is used to divide the pixels on each segment mask into two groups.

7.3 Classification and Visualisation Results

In this section, we evaluate the effectiveness of our proposed method using the MINI-RGBD dataset. We first compare the performance of our method on fidgety movement detection with our baseline methods. Next, we present the visualization results as qualitative analysis in Section 7.3. We follow the standard protocol as in [165, 166, 153] to conduct a leave-one-subject out cross-validation to ensure the results presented in this section are obtained based on *unseen data* during the training process.

Quantitative Evaluation on the Fidgety Movement Detection Results

To demonstrate the overall performance of our proposed framework, we first evaluate the CP prediction of the whole input video as explained in Section 7.2. We compare our results with the existing methods using additional performance metrics, and present this in Table 7.1. Using our framework, we achieved a perfect prediction with 100% accuracy, which highlights the effectiveness of our proposed framework over the previous work ([165, 166, 153]). In this instance

Table 7.1: Classification accuracy comparison between our proposed visualisation framework and the selected baseline methods.

Method	Accuracy	Sensitivity	Specificity
[165] w/ LDA	66.7%	50.0%	75.0%
[165] w/ SVM	83.3%	50.0%	**100.0%**
[165] w/ Decision Tree	75.0%	50.0%	87.5%
[165] w/ kNN (k=1)	75.0%	25.0%	**100.0%**
[165] w/ kNN (k=3)	50.0%	00.0%	75.0%
[165] w/ Ensemble	66.7%	50.0%	75.0%
FCNet [166]	83.3%	75.0%	87.5%
Conv1D-1 [166]	83.3%	75.0%	87.5%
Conv1D-2 [166]	91.7%	75.0%	**100.0%**
Conv2D-1 [166]	83.3%	75.0%	87.5%
Conv2D-2 [166]	83.3%	75.0%	87.5%
Movement Complexity Index [153]	91.7%	**100.0%**	87.5%
Our method	**100.0%**	**100.0%**	**100.0%**

we were able to outperform both the results from our initial feasibility study and our previously proposed deep learning methods across the three selected performance metrics. We note that the sensitivity metric is particularly important in this case, since it represents the percentage of correctly identified positive cases i.e. cases of those with absent FMs and subsequently at higher risk of developing CP. Given that this is the minority class within the dataset we find that our perfect classification performance here is particularly encouraging.

Qualitative Evaluation on the Visualization Results

We further provide qualitative results to demonstrate the effectiveness of our proposed framework. As presented in Section 7.1, we detect the absence (FM-) or presence (FM+) of FMs for each body part in each temporal segment. The body parts with a prediction of FM- will be highlighted in red, an example of which is illustrated in Figure 7.3. From the results, it can be seen that the highlighted body-parts generally show less complex or more repetitive movements in the videos annotated as FM-. As shown in Figure 7.3 (a), the arms are showing a lack of movement and are subsequently predicted as FM- using our framework. Additionally, the legs are predicted as FM+ as they are correspondingly showing some movements in that temporal segment. Figure 7.3 (c) shows an example with monotonous arm and leg movements and our method highlights those body parts as FM- accordingly. For the videos annotated as FM+, such as the example shown in Figure 7.3 (b), it can be seen that a much wider variety of movements can be observed. We suggest that our

(a) Positive (abnormal) example

(b) Negative (normal) example

(c) Positive (abnormal) example

Figure 7.3: Examples of the visualization module. Body-parts where FMs are absent are highlighted in red.

proposed visualization method provides effective visual feedback to the user, as such clinicians can pay greater attention to the highlighted segments for further analysis.

7.4 Concluding Remarks

In this work, we present a new framework for detecting fidgety movements of infants spatiotemporally using the pose-based features previously established and extracted from 2D RGB videos. Experimental results demonstrated that the new method not only achieves perfect prediction, with 100% accuracy, but also provides the user with visualization on how the machine-learning based framework made the overall prediction relating to the abnormality of the infant's movement.

Our proposed approach also enables further exploration of the temporal aspects of the GMA through the extraction of additional spatio-temporal features for further analysis and improved interpretability.

However, whilst our system is able to provide rudimentary visual feedback to the user, additional visualization tools would be useful to exploit the extracted spatiotemporal information, and provide additional predictive aid to clinicians for this complex diagnostic task. We intend to further implement our method on the data gathered in a real-world clinical setting, as well as experiment with our other proposed features and explore additional GMA relevant features and visualisation methods to provide meaningful and explainable feedback to the user.

Chapter 8

Conclusion and Future Work

In this chapter, we provide a retrospective overview of the project. We touch upon the motivation for our work and the related works, and discuss our proposed methods, the results of our experiments, and the overall conclusions we have drawn from the project. Finally we present some ideas relating to possible future work in this field.

8.1 Conclusion

CP is an umbrella term used to describe a group of lifelong neurological conditions, attributed to non-progressive damage to the brain in early infancy, which cause movement difficulties. These movement difficulties typically affect mobility, posture and coordination, but can also cause problems with speech articulation, swallowing, vision, and can contribute towards a reduced ability to learn new skills. In order to provide opportunities for the best possible outcome for an infant's development, early diagnosis of CP is considered essential.

We examined several prominent methods for the diagnosis of CP using both physical examination and neurological imaging. We established that the challenges of applying these assessments relate to the availability of appropriately trained and skilled clinicians, but that one of the leading methods of diagnosis was the GMA, and that it was feasible that this test could be adopted to form the basis of an automated prediction tool using machine learning techniques.

Whilst several approaches for automated assessment exist, we identified that these methods typically faced several challenges based upon the techniques used. As such, we suggest that there is a requirement for the adoption of new state-of-the-art methods, which make use of advances in the fields of computer-vision and human activity recognition, in this domain.

We suggest that pose-estimation, image segmentation and histogram representations offer significant benefits over other methods in both the analysis of human motion and interpretability of the associated data. As such, we present several new pose-based histogram features which retain human interpretability and are directly mapped to the assessment criteria in the clinical GMA guidelines and HINE Optimality Score. We also present the need for unbiased evaluation of both our proposed frameworks and existing methods using shared datasets, and the importance of real-world clinical data in this evaluation.

In our first series of experiments, we established and evaluated two new pose based features, namely HOJO2D and HOJD2D. We determined the viability of this method and concluded that pose-based assessment is a feasible alternative to other methods from the literature, providing suitable information to allow for classification. We also evaluated the strengths of body-part based assessment and determined that the most discriminative body parts in this setting were the limbs. We concluded that pose-based assessment is well suited to the task of infant motion analysis, re-

porting state-of-the-art performance of 91.67% accuracy on the MINI-RGBD dataset, when using several combinations of fused (i.e. concatenated) features.

In our next work, we extended and enhanced the previous feature extraction and classification pipeline to incorporate deep learning frameworks. In this study we wanted to assess the possibility of using deep learning frameworks simply to classify the extracted hand-crafted features. In doing so we are able to retain the interpretability of the framework through the use of feature engineering, but also capitalise upon the potential benefits of improved classification performance brought about by the adoption of deep learning. Our experimental results showed that the application of deep learning in this domain is viable, providing a robust performance of 91.67% across varying feature sets. However, we did not achieve class leading performance in all areas, which may have been due to the limited size of the synthetic dataset, but this is somewhat offset by the improved consistency demonstrated throughout. Our ablation tests support these observations, and we suggest that with a larger dataset greater performance improvements may also be evident.

In our next and perhaps most comprehensive work, we proposed several new and improved interpretable pose-based features, namely: Histograms of Angular Displacement (HOAD2D), Histograms of Relative Joint Orientation (HORJO2D), Histograms of Relative Joint Angular Displacement (HORJAD2D), Fast Fourier Transform of Joint Orientation (FFT-JO), Fast Fourier Transform of Joint Displacement (FFT-JD), Histograms of Joint Orientation (HOJO2D), and Histograms of Joint Displacement (HOJD2D). We also reimplemented several methods from the literature to serve as baselines for comparison using shared datasets. Notably, in this work we also made use of our newly constructed RVI-38 dataset, which consists of real-world clinical data. This allowed us to assess the robustness of each method on two separate datasets for a more representative comparison. Our extensive evaluation, using a full suite of performance metrics and statistical analytics, suggests that the performance of our individual new features generally outperforms the previous methods. However, based upon our previous work, our motivation was also to evaluate the performance of our features when fused, since this should provide greater generalisability and more accurately represent the complexity associated with the GMA. Our findings suggest that our fused features provide state-of-the-art results on both the MINI-RGBD dataset (F1: 100%, MCC 100.00%) and the RVI-38 dataset (F1: 90.91%, MCC: 89.89%), which highlights their improved performance and greater robustness over all previous methods.

In our final work, we presented a new framework which was capable of detecting FMs of infants spatiotemporally. This meant that we were able to augment our classification pipeline with visualisation data in order to further improve the interpretability. We again used the same features extracted previously but in this case we segmented the motion sequence and annotated each segment allowing for us to evaluate risk related movements spatiotemporally. We evaluated our framework using standard metrics and again achieved a state-of-the-art performance of 100% accuracy on the MINI-RGBD dataset. However in contrast with each of the previously proposed methods, we were also able to qualitatively evaluate a visualisation component within the framework, which allowed us to visualise and evaluate specific risk related movements, subsequently providing meaningful feedback to clinicians.

We suggest that by utilising pose-based features, we make the likelihood of collaborative working within the healthcare domain more viable, due to the inherently anonymised and unidentifiable patient data. As such, we have made the pose data, the GMA based dataset annotations, and the feature extraction and classification code for each of our frameworks, available to the community in an effort to help stimulate research in this impactful area. In the next section we discuss our potential future works and how these might lead to further improvements in this field.

8.2 Future Work

In this thesis we have explored several methods for the prediction of CP based upon infant movements. Whilst the results we have achieved thus far are encouraging, there are several areas which we would like to address in future works.

Firstly, we hope to explore further improving the interpretability of CP prediction models by allowing clinicians to form a more significant part of the feedback loop. We would like to extend the implementation of our proposed features by producing additional visualisations, which could include input from clinicians. This would also potentially allow us to identify additional spatio-temporal features with a higher degree of accuracy through improved annotation, meaning that we are able to extract more meaningful features from the data through approaches such as *active learning* or *weakly supervised learning*, allowing further exploration of the temporal aspect of physical assessments.

Additionally, we would like to gather more video data from clinical settings, and additional annotations relating to FMs. This could allow us more opportunities to experiment with other options, and to minimise model overfitting problems, to help consolidate our results, and to improve the data imbalance. Our goal is to improve the overall sensitivity of the proposed frameworks such that we are better able identify to the positive samples, which would be essential as part of a screening programme. As such we may also explore data augmentation to help deal with any class imbalance and further enhance the credibility of our results.

Furthermore, given that our method relies upon the OpenPose framework [151], which is trained using adult data, we would like to investigate enhancing this framework through domain adaptation or the generation of synthetic training data to further improve the raw pose output, which would help to address any potential inconsistencies in the data. Although the results demonstrated suggest that these inconsistencies are largely dealt with by our qualitative assessment and automated pre-processing techniques, we would still like to enhance the pose-estimation output to make the framework more specific to infant body dimensions and posture.

Also, we would like to explore the possibility of implementing some of our findings in other projects into the pipelines proposed in this work. Specifically, we would like to adapt our work on assessing facial symmetry using augmented reality [167] as a means of tracking the body asymmetry associated with hemiplegic CP. We would also like to investigate the 2D to 3D prior-less reconstruction method proposed in our work [168] to generate additional motion data in three dimensions, potentially allowing us to assess the infant movements even more comprehensively.

Finally, given that there are several tests available for the prediction of CP, we may evaluate these individually using specific annotation generated using each method. In doing so we may be able to quantify the features associated with each of these tests and subsequently combine several tests to generate a more holistic understanding of the infant body movements and neurological development. Whilst we did this in a limited manner in our final work, we would like to further expand upon this. It could therefore be conceivable that a combined test such as this would provide greater robustness and generalisability than existing methods, and potentially offer greater insight into infant neurodevelopment.

Appendix A

Ethical Approval Documents

- Research Protocol
- Parental Consent Covering Letter
- Parent Information Sheet
- Consent Form

Research Protocol

Sensing Movement using Action Recognition Technology in Babies

Investigators

- Professor Nicholas Embleton, Consultant Neonatal Paediatrician, Newcastle Neonatal Service, Newcastle upon Tyne Hospitals NHS Foundation Trust (Principle Investigator)
- Ms Claire Marcroft, Senior Paediatric Physiotherapist, Newcastle Neonatal Service, Newcastle upon Tyne Hospitals NHS Foundation Trust
- Mr Kevin McCay, PhD Student, Northumbria University
- Dr Edmond Ho, Senior Lecturer, Northumbria University

Background

The automated recognition of human activity has wide ranging applications and is widely used in healthcare. This project proposes that this technology could be used in paediatric practice to aid with early diagnosis of movement disorder such as cerebral palsy in infancy.

Making an early diagnosis of cerebral palsy or other movement disorders is important but current methods lack sensitivity and specificity and are time consuming. Applying current tools and assessments is dependent upon the availability of fully trained practitioners - these skills take many years to develop. However, all tools are susceptible to observer fatigue, contain a degree of personal subjectivity and are reliant upon a suitable behavioural state of the infant. If infants are upset or too tired when they attend clinic then the assessment may have to be stopped or repeated.

There is scope to improve the accuracy and accessibility of existing tests through computer-based evaluations; these evaluations could also conceivably provide quantifiable evidence to clinicians. The development of automated systems could also help to significantly reduce the time and subsequent cost associated with these current diagnostic practices. We want to try and develop techniques that will use non-invasive film/video assessments of normal movement in babies at rest. Our pilot work will use existing film clips that were collected during our routine clinic visits.

Study Design

Observational study.

Sample

Up to 200 preterm infants born <30 weeks gestational age (GA) or <1000g birth weight.

Method for Study

Between 35 weeks GA and 5 months post term corrected age, in the routine neonatal developmental clinic, infants undergo video assessment (GM assessment). The video assessment is undertaken by a clinician trained experienced in the GM Assessment and involve the infant's movements being videoed for up to 15 minutes when the infant is in an awake and alert state.

The video recording is observed and categorised based on the quality of the general movements observed using Gestalt perception. The first step is to differentiate between normal GMs and abnormal GMs. Secondly, further sub categorisation of abnormal GMs such as 'abnormal fidgety movements', 'absence of fidgety movements' and 'cramped synchronised GMs' is undertaken.

Members of the direct healthcare team who provide follow up for these infants (Dr Embleton and Ms Claire Marcroft, Physio) will identify infants whom we have existing film footage of and will screen clips for suitability. Dr Embleton will write to parents and include an information sheet and consent form. Parents will return the consent form by post but will be given an opportunity to meet with Dr Embleton and team in person at the RVI if they wish. We are also asking parents who are attending routine developmental clinics, if we can share the video gathered as part of routine clinical care for use in this study, an information sheet and consent form will again be provided.

The video footage will then be securely transferred to Dr Edmond Ho and Kevin McCay at Northumbria University for use in training, testing and validating a system to help automate the identification of infants at risk of developing movement disorders.

Study Population

The study group will be recruited from the Special Care Baby Unit at the Royal Victoria Infirmary, Newcastle. Infants who fit the inclusion/exclusion criteria will be invited to participate

Inclusion Criteria

- Preterm infants born between 23-29 weeks gestation and aged between 12 and 20 weeks corrected age at the time of assessment.
- Parents with adequate understanding of English to ensure informed consent.

Exclusion Criteria

- Infants with a suspected or diagnosed genetic abnormality
- Infants considered medically unstable by attending medical staff (e.g. frequent apnoea, poor thermoregulation)
- parents who are incapable of giving informed consent

Data Analysis

We will be attaining visual features automatically through a computer vision framework. These features will then be mapped to the existing manual assessment schemes. The accuracy of the resultant prediction provided by the framework will be quantitatively compared with the existing assessment scheme and expert analysis. The intention is that this could also provide quantitative evidence of a standardised assessment scheme. We anticipate measurably improved accuracy in automated pose estimation and action recognition for infants.

Ethical Considerations

Ethical approval will be obtained from National Research Ethics Committee Service and Newcastle upon Tyne Hospitals NHS Foundation Trust Joint Research Office prior to commencing the study. Caldicott Approval will also be obtained.

In this pilot study we will be using existing video data, of infants born preterm, from Ward 35 (SCBU), Level 4, Leazes Wing, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne, NE1 4LP.

Data used after parental consent and will be anonymised by masking faces/body parts or only including the computer-generated skeleton in any external work. No identifiable clinical information will be shared outside of clinical records. Personal data will remain in the NHS record with the subject identified by a unique study code. Data will be analysed in a code-linked, anonymised fashion by the team at Northumbria University (Dr Edmond Ho and Kevin McCay) and the team at the RVI in Newcastle (Dr Nick Embleton and Claire Marcroft).

The data will be analysed both within the secure premises of the RVI and Ellison Building at Northumbria University. Data will be transferred from the internal NHS system at the RVI to a hardware encrypted portable hard drive at a specified time. This hard drive will then be taken directly to the secure computing facility at Northumbria University where the data will be transferred to a local drive on a secure, un-networked PC. The data contained on the portable hard drive will then be destroyed.

We do not consider that the study raises any important ethical issues over and above the need to follow information governance issues. We will ask parents for consent to use existing footage and do not think this will be burdensome or upset parents in any way.

Ethical approval will be obtained from a local research ethics committee prior to commencing the study. Approval will also be sought from the Newcastle Joint research Office on behalf of the Newcastle Hospitals NHS Foundation Trust. We do not consider there to be any risk to the infants taking part.

Informed Consent

The study will be explained in detail to the person(s) with parental responsibility for the infant. A comprehensive participant information leaflet will be provided. Written consent for the infant's video assessments to be included in this study must be granted by the person(s) with parental responsibility. The person(s) with parental responsibility will have the opportunity to remove consent at any time without giving reason.

Data Management

All clinical data obtained during the video capture sessions has be documented and stored in electronically on a networked Newcastle upon Tyne Hospitals NHS Foundation Trust (NUTH) drive. This drive is automatically backed-up and only accessible by NUTH login.

No identifiable clinical information will be shared outside of clinical records. Personal data will remain in the NHS record, participants will be assigned a unique study number and this log will be kept in the investigator site file. The computer research team have no access to personal data.

The data will be analysed both within the secure premises of the RVI and Ellison Building at Northumbria University. Data will be shared with the research team in Northumbria University under the supervision of the PI (Dr Nick Embleton) and subsequently will be included in publication - in an anonymised and non-identifiable form. This is included in the patient information sheet and on the consent form. SOPs for the NHS Trust will apply, and the data protection act principles will be adhered to at all times.

Address Line 1
Address Line 2
Address Line 3
Postcode

Date: 06/06/2019

Address Line 1
Address Line 2
Address Line 3
Postcode

Dear Parent/Guardian,

We are inviting babies and their families who have received care in the Special Care Baby Unit at the Royal Victoria Infirmary in Newcastle to be involved in a clinical research study. We hope this research will help improve our care in the future.

As part of your baby's routine clinical care, you will probably remember that the physio team took video recordings of them moving. This may have been done either on the neonatal unit or during one of your developmental follow-up appointments. We use these videos to help us understand how your baby's movements develop over time.

We are writing to ask for your consent to use the videos we have already taken of your baby to help us learn more about how babies born prematurely develop their movement skills. We are working with computer scientists at Northumbria University who are trying to develop computer software to help us better assess babies movement.

If you would like to know more about the project, please read through the enclosed information sheets. If you have any queries relating to this study, please feel free to contact any of the members of the research team for more information.

If you have read the information sheets and would like to help us with our study, please fill out and return the enclosed consent form using the stamped self-addressed envelope. Thank you for helping with this important research.

Kind regards,

PARENT INFORMATION SHEET

Sensing Movement using Action Recognition Technology in Babies
IRAS ID Number: 252317

We would like to invite your baby to take part in a research study. Before you decide you need to understand why the research is being done and what it would involve for you.

- Part 1 tells you the purpose of the study and what will happen if you choose to take part.
- Part 2 gives you more detailed information about the conduct of the study.

Contact Information

Please ask if you would like more information or if anything is not clear.

Professor Nick Embleton
Consultant Neonatal Paediatrician – Royal Victoria Infirmary, Newcastle upon Tyne
Email: nicholas.embleton@newcastle.ac.uk

Claire Marcroft
Academic Physiotherapist – Royal Victoria Infirmary, Newcastle upon Tyne
Email: claire.marcroft@newcastle.ac.uk

Kevin McCay
PhD Student – Northumbria University, Newcastle upon Tyne
Email: kevin.d.mccay@northumbria.ac.uk

Edmond Ho
Senior Lecturer – Northumbria University, Newcastle upon Tyne
e.ho@northumbria.ac.uk

If you have any additional queries or complaints that have not been dealt with by the above research team, please contact the Northumbria University Research Policy Manager Laura Hutchinson by email at laura.hutchinson2@northumbria.ac.uk.

PART 1

Research Overview

Babies who are born prematurely have an increased risk of developmental difficulties including the way they move. It is important that healthcare professionals can identify which babies may go on to have difficulties, so they can offer support to parents and access therapy services (such as physiotherapy). Additionally, it is also important that professionals can reassure parents of babies who are progressing without difficulties.

We are computer scientists based at Northumbria University, Newcastle and hope that by working closely with healthcare professionals (physiotherapists and doctors) based at the Royal Victoria Infirmary in Newcastle, we can use state of the art human motion analysis to develop new ways to identify which babies are most at risk.

What will be happen if my baby takes part?

The physiotherapists in Newcastle routinely use video recordings to monitor the developmental progress for all babies born < 30 weeks gestation. This is part of routine care. They set up a video camera on a tripod at the end of a cot space in the neonatal unit or on a comfy mat on the floor during a clinic appointment and record your baby's spontaneous movements for up to 3 minutes. They also collect details of your baby's gestational age, weight, sex and corrected age (this is the age of your baby calculated from their due date).

Why has my baby been invited to take part?

We are inviting infants, from the Special Care Baby Unit at the Royal Victoria Infirmary in Newcastle, who have been born prematurely (under 30 weeks gestation) to be involved. We are writing to the parents of infants who have already had video recordings taken as part of routine clinical care, to ask permission to share the recordings for use in this study. We are also asking parents who are attending routine developmental clinics, if we can share the video gathered as part of routine clinical care for use in this study.

If you think you would like your baby to take part, you will be asked to sign a consent form. If you choose to take part**, there will be NO change to the routine care your baby will receive.** We will use the video recordings of your baby taken during the routine physiotherapy assessments.

Does my baby have to take part?

No. It is up to you to decide. If you do decide to participate, we will then ask you to sign a consent form to show you have agreed to take part. Whether you choose to participate or decide not to take part will not affect the standard of care your baby will receive in the NHS at any time.

What are the possible benefits of taking part?

There is no direct benefit to your baby from taking part. The study will help us to gather information about the effectiveness of the assessments we routinely use. In the longer term it might help us look after babies more effectively.

PART 2

Time Commitment

The recording session typically takes **15** minutes, but no additional recording will be required at this stage.

Participant's rights

You may decide to stop being a part of the research study at any time without explanation. You have the right to access the information gathered at any time. You have the right to ask that any data you have supplied to that point be withdrawn and/or destroyed. If you have any questions you should ask the researcher before the study begins. If you wish to contact the research team at any time, please use email the addresses provided above.

There is no risk to your baby taking part as **all data gathered is collected as part of routine care**. However, if you have a concern about any aspect of this study, you should ask to speak to a member of the research team who will do their best to answer your questions (please see contact information).

What will happen to the data?

All data will be treated confidentially and will be stored securely on an encrypted, password protected computer. The data will only be accessible to those directly associated with the research and approved by all relevant ethical approval bodies. If the data is to be used directly in published works (such as academic journals), it will be fully anonymised prior to publication. Data collected can be anonymised by masking faces/body parts or only including images of the computer-generated skeleton. If a request is received to have data withdrawn and destroyed it will be deleted from the secure device immediately.

How your information may be used

If you agree to take part in this research study, we will collect the minimum personally-identifiable information needed for the purposes of the research project. Information about your baby will be used in the ways needed to conduct and analyse the research. NHS organisations may keep a copy of the information collected. Depending on the needs of the study, the information may include personal data that could identify you. You can find out more about the use of patient information for the study from the research team.

Your baby's personal information (initials, date of birth, contact details and your consent form) will be assigned a unique identification number and all data collected will be linked to this number. This will ensure that your details remain anonymous as personal details will not be transferred outside of the RVI i.e. the computer science team will not know your baby's name, date of birth, address etc.

The video recording taken in developmental clinic will also be kept on a password protected Newcastle Upon Tyne Hospital Network networked computer. The video transferred to Northumbria University will be kept on a password protected networked computer. Only study team members will be able to access the video.

You may be asked to provide information about your child's health to the research team, for example in a questionnaire. Sometimes information about you will be collected for research at the same time as for your clinical care. In other cases, information may be copied from your health records. Information from your health records may be linked to information from other places such as central NHS records, or information about you collected by other organisations. You will be told about this when you agree to take part in the study.

Will any information be stored?

Yes, but only with your permission.

What will happen to the results of the research study?

It is intended that the results will be published in medical/scientific journals and presented at medical/scientific meetings. Your baby will not be identified in any of the reports or publications. A summary of the findings of the study will be available to you at the end of the project.

Keeping information for future research

Information about your baby that is collected during a research study may be kept securely to be used in future research in any disease area, including research looking at social and economic factors affecting health. This may include combining it with information about you held by other health or government organisations such as NHS Digital. Usually the information is combined by matching information that has the same NHS number. Doing this makes maximum use of the information you have provided and allows researchers to discover more.

Researchers may not be able to specify all the possible future uses of the information they keep. It could include providing the information to other researchers from NHS organisations, universities or companies developing new treatments or care. Wherever this happens it will be done under strict legal agreements. The information about you will be depersonalised wherever possible so that you cannot be identified. Where there is a risk that you can be identified your data will only be used in research that has been independently reviewed by an ethics committee.

On rare occasions NHS organisations may provide researchers with confidential patient information from your health records when we are not able to seek your agreement to take part in the study, for example because the number of patients involved is too large or the NHS organisation no longer has your contact details. Researchers must have special approval before they can do this.

Your choices about health and care research

If you are asked about taking part in research, usually someone in the care team looking after you will contact you. People in your care team may look at your health records to check whether you are suitable to take part in a research study, before asking you whether you are interested or sending you a letter on behalf of the researcher.

It's important for you to be aware that if you (or your baby) are taking part in research, or information about you is used for research, your rights to access, change or move information about you are limited. This is because researchers need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from a study, the sponsor will keep the information about you that it has already obtained. They may also keep information from research indefinitely.

If you would like to find out more about why and how patient data is used in research, please visit the Understanding Patient Data website https://understandingpatientdata.org.uk/what-you-need-know.

Who has reviewed the study?

All research in the NHS is looked at by an independent group of people called a Research Ethics Committee to protect your safety, rights, wellbeing and dignity. This project has been reviewed by the Bromley Research Ethics Committee.

Thank you for reading this information sheet and considering taking part. If you wish to participate then you can discuss this with a member of the research or clinical team and they will help to you proceed to signing the consent form. Or you can return the consent form in the envelope provided after you have had time to consider it.

Northumbria University NEWCASTLE

The Newcastle Upon Tyne Hospitals NHS Foundation Trust

CONSENT FORM

Project Title: Sensing Movement using Action Recognition Technology in Babies

IRAS ID Number: 252317

Principal Investigators: Professor Nick Embleton
Claire Marcroft
Kevin McCay

Please initial in each box if you agree with the following statements:

I have carefully read and understood the Parent Information Sheet dated…….. (version ……) which contains important safety, privacy and other information about the study.	
I have had an opportunity to consider the information presented to me, ask questions, discuss the study and I have received satisfactory answers.	
I understand that participation is entirely voluntary and that I am free to withdraw from the study at any time, without having to give a reason for withdrawing, and without prejudice.	
I understand that no payment will be received for participation in the study.	
I understand that all information collected will be held in confidence, and that if presented, identities will not be disclosed (i.e. data will be codified).	
I hereby confirm that I give consent for recordings captured in previous sessions to be made available to the research team as discussed in the Parent Information Sheet.	
I understand that relevant sections of my baby's medical notes, test results and other routinely-collected electronic health records may be looked at by my doctor and the associated care team. I give permission for relevant information to be provided to collaborating researchers once all identifiable information is hidden (codified).	
I agree to take part in the study.	

Please initial in each box if you agree with the following Northumbria University clauses:

Clause A: I understand that other individuals may see the recording(s) and be asked to provide ratings/judgments. The outcome of such ratings/judgments will not be conveyed to me. My name or other personal information will never be associated with the recording(s).	

Clause B: I understand that the recording(s) may also be used for teaching/research purposes and may be presented to students/researchers in an educational/research context. My name or other personal information will never be associated with the recording(s).	
Clause C: I understand that the recording(s) may be published in an appropriate journal/textbook or on an appropriate Northumbria University webpage, which would automatically mean that the recordings would potentially be available worldwide. My name or other personal information will never be associated with the recording(s). I understand that I have the right to withdraw consent at any time prior to publication, but that once the recording(s) are in the public domain there may be no opportunity for the effective withdrawal of consent	
Clause D: I also consent to the retention of this data under the condition that any subsequent use also be restricted to research projects that have gained ethical approval from Northumbria University.	

Name of the parent/guardian of the participant

.....

Signature of the parent/guardian of the participant

..... Date.....

Name of researcher taking consent

.....

Signature of researcher taking consent

..... Date.....

Reference Number

.....

Acronyms

AC	Accuracy 86
AI	Artificial Intelligence viii, 24, 27, 31, 54, 79
AMD	Absolute Motion Distance 67, 75, 98, 103
BINS	Bayley Infant Neurodevelopmental Screener 12, 18
CNN	Convolutional Neural Network 31, 43, 59
CoM	Centroid of Motion 41, 43, 66, 75
CP	Cerebral Palsy iii, 1–3, 6, 7, 9–24, 29, 31, 34, 36, 37, 41, 43, 50, 52, 65, 67–69, 77, 78, 107,
CPP	Cerebral Palsy Predictor 67, 75
CT	Computerised Tomography 12, 19–21
DT	Decision Tree 29, 85
EEG	Electroencephalogram 12, 19, 21
ENS	Ensemble 30, 79, 85, 87–89, 100, 107, 114
F1	F1 Score 86, 98–100, 121
FFT	Fast Fourier Transform 68, 72, 73
FFT-JD	Fast Fourier Transform of Joint Displacement 72–75, 84, 97–99, 103, 106, 107, 121
FFT-JO	Fast Fourier Transform of Joint Orientation 73–75, 84, 97–100, 103, 106, 107, 121
FM+	Fidgety Movements Present 57, 58, 105–107, 112–114, 116
FM-	Fidgety Movements Absent 57, 58, 105–107, 112–114, 116

FMs	Fidgety Movements 6, 13, 14, 68, 112–114, 116, 117, 122, 123
FN	False Negative 86, 87
FP	False Positive 86, 87
FPS	Frames Per Second 61
GMA	General Movements Assessment iii, 2–5, 7, 12–15, 24, 37–39, 41, 53, 56–58, 60, 65, 69, 72, 78, 107, 118,
GMs	General Movements 13, 14, 34, 73
HAI	Hand Assessment for Infants 2, 12, 17
HINE	Hammersmith Infant Neurological Examination 15–17, 108, 120
HNNE	Hammersmith Neonatal Neurological Examination 15
HOAD2D	Histograms of Angular Displacement 70–72, 74, 75, 84, 97–99, 103, 105, 121
HOJD2D	Histograms of Joint Displacement ix, 73–75, 78–80, 84, 87–89, 91, 97–99, 103, 107, 113, 120, 121
HOJO2D	Histograms of Joint Orientation ix, 73–75, 78–80, 84, 87–90, 97–99, 103, 107, 113, 120, 121
HORJAD2D	Histograms of Relative Joint Angular Displacement 72, 74, 75, 84, 97–99, 103, 105, 106, 121
HORJO2D	Histograms of Relative Joint Orientation 71, 72, 74, 75, 84, 97–99, 103, 105, 106, 121
kNN	K-Nearest Neighbour 30, 79, 85
LAPI	Lacey Assessment of Preterm Infants 2, 12, 15
LDA	Linear Discriminant Analysis 30, 79, 85, 88
LDOF	Large Displacement Optical Flow 41, 42
LR	Logistic Regression 29, 85
MCC	Matthews Correlation Coefficient 87, 98–100, 121
MCI	Movement Complexity Index 68, 75
MINI-RGBD	Moving INfants In RGB-D 57, 87, 90, 96, 98–101, 103–106, 115, 121, 122
MRI	Magnetic Resonance Imaging 3, 12, 19–21
MWC	Magnitude of Wavelet Coefficients 67, 68, 75, 98, 103

PCA	Principal Component Analysis 30
PR	Precision 86
QoM	Quantity of Motion 66, 75
RE	Recall 86
RF	Relative Frequency 67, 75, 103
RVI-38	Royal Victoria Infirmary-38 Dataset iii, 6, 96, 98–101, 103–106, 108, 121
RVI-GMA	Royal Victoria Infirmary - General Movements Assessment 57
SATCo	Segmental Assessment of Trunk Control 12, 19, 42, 43
SE	Sensitivity 86
SP	Specificity 86
SVM	Support Vector Machine 29, 37, 85
TN	True Negative 86, 87
TP	True Positive 86, 87
WMs	Writhing Movements 13

References

[1] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14(2):201–211, 1973.

[2] J.K. Aggarwal and Lu Xia. Human activity recognition from 3d data a review. *Pattern Recognition Letters*, 48:70 – 80, 2014.

[3] Li Yao, Yunjian Liu, and Shihui Huang. Spatio-temporal information for human action recognition. *EURASIP Journal on Image and Video Processing*, 2016(1):39, Nov 2016.

[4] G Sinha, R Shahi, and M Shankar. Human Computer Interaction. In *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, pages 1–4. IEEE, nov 2010.

[5] Scope UK. Cerebral palsy (CP), 2020.

[6] Claire Marcroft, Hons Mcsp, Patricia Dulson, Richard Hearn Mbchb, Anna Basu, Bmbch Ma, Nicholas Embleton, and Hons Mbbs. Does the Lacey Assessment of Preterm Infants predict cerebral palsy in extremely preterm infants ? A pilot study. *Association of Paediatric Chartered Physiotherapists*, 5(2), 2014.

[7] Mijna Hadders-Algra. Early diagnosis and early intervention in cerebral palsy. *Frontiers in Neurology*, 5(SEP):1–13, 2014.

[8] Mindy Lipson Aisen, Danielle Kerkovich, Joelle Mast, Sara Mulroy, Tishya A.L. Wren, Robert M. Kay, and Susan A. Rethlefsen. Cerebral palsy: Clinical care and neurological rehabilitation. *The Lancet Neurology*, 10(9):844–852, 2011.

[9] Anna Purna Basu and Gavin Clowry. Improving outcomes in cerebral palsy with early intervention: New translational approaches, 2015.

[10] Claire Marcroft, Aftab Khan, Nicholas D. Embleton, Michael Trenell, and Thomas Plötz. Movement recognition technology as a method of assessing spontaneous general movements in high risk infants. *Frontiers in Neurology*, 6(JAN):284, 2015.

[11] Joy Olsen, Peter Marschik, and Alicia Spittle. Do fidgety general movements predict cerebral palsy and cognitive outcome in clinical follow-up of very preterm infants?, 2018.

[12] Nathalie Maitre. Skepticism, cerebral palsy, and the General Movements Assessment. *Developmental Medicine and Child Neurology*, 2018.

[13] Margot Bosanquet, Lisa Copeland, Robert Ware, and Roslyn Boyd. A systematic review of tests to predict cerebral palsy in young children. *Developmental Medicine & Child Neurology*, 55(5):418–426, 2013.

[14] Iona Novak. Evidence-based diagnosis, health care, and rehabilitation for children with cerebral palsy. *Journal of Child Neurology*, 29(8):1141–1156, 2014.

[15] Hao Shu Fang, Shuqin Xie, Yu Wing Tai, and Cewu Lu. RMPE: Regional Multi-person Pose Estimation. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-Octob:2353–2362, 2017.

[16] R. A. Güler, N. Neverova, and I. Kokkinos. Densepose: Dense human pose estimation in the wild. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7297–7306, June 2018.

[17] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *CVPR*, 2017.

[18] Supasorn Suwajanakorn, Noah Snavely, Jonathan Tompson, and Mohammad Norouzi. Discovery of latent 3D keypoints via end-to-end geometric reasoning. *Advances in Neural Information Processing Systems*, 2018-December(1):2059–2070, 2018.

[19] Lu Yang, Qing Song, Zhihui Wang, and Ming Jiang. Parsing r-cnn for instance-level human analysis, 2018.

[20] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid. Learning from synthetic humans. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4627–4635, July 2017.

[21] Zehua Sun, Jun Liu, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, and Gang Wang. Human action recognition from various data modalities: A review. *CoRR*, abs/2012.11866, 2020.

[22] Yucheng Chen, Yingli Tian, and Mingyi He. Monocular human pose estimation: A survey of deep learning-based methods. *Computer Vision and Image Understanding*, 192(December 2019):102897, 2020.

[23] Sara Moccia, Lucia Migliorelli, Virgilio Carnielli, and Emanuele Frontoni. Preterm Infants' Pose Estimation With Spatio-Temporal Features. *IEEE Transactions on Biomedical Engineering*, 67(8):2370–2380, 2020.

[24] Claire Chambers, Nidhi Seethapathi, Rachit Saluja, Helen Loeb, Samuel R. Pierce, Daniel K. Bogen, Laura Prosser, Michelle J. Johnson, and Konrad P. Kording. Computer vision to automatically assess infant neuromotor risk. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(11):2431–2442, nov 2020.

[25] Kevin Lin, Lijuan Wang, Kun Luo, Yinpeng Chen, Zicheng Liu, and Ming-Ting Sun. Cross-domain complementary learning with synthetic data for multi-person part segmentation. *arXiv preprint arXiv:1907.05193*, 2019.

[26] Martin Bax, Murray Goldstein, Peter Rosenbaum, Alan Leviton, Nigel Paneth, Bernard Dan, Bo Jacobsson, and Diane Damiano. Proposed definition and classification of cerebral palsy, april 2005. *Developmental Medicine and Child Neurology*, 47(8):571–576, 2005.

[27] NICE. NICE seeks to improve diagnosis and treatment of cerebral palsy, Jan 2017.

[28] Chitra Sankar and Nandini Mundkur. Cerebral Palsy - definition, classification, etiology and early diagnosis. *Indian Journal of Pediatric s*, 72(10):865–868, 2005.

[29] Peter Rosenbaum, Nigel Paneth, Alan Leviton, Murray Goldstein, Martin Bax, Diane Damiano, Bernard Dan, and Bo Jacobsson. A report: The definition and classification of cerebral

palsy april 2006. *Developmental medicine and child neurology. Supplement*, 109:8–14, 03 2007.

[30] NHS. NHS - cerebral palsy overview, Feb 2020.

[31] Maryam Oskoui, Franzina Coutinho, Jonathan Dykeman, Nathalie Jette, and Tamara Pringsheim. An update on the prevalence of cerebral palsy: A systematic review and meta-analysis. *Developmental medicine and child neurology*, 55, 01 2013.

[32] Susan M Reid, Elaine Meehan, Sarah McIntyre, Shona Goldsmith, Nadia Badawi, Dinah S Reddihough, and AusACPDM. Temporal trends in cerebral palsy by impairment severity and birth gestation. *Developmental Medicine & Child Neurology*, 58(S2):25–35, 2016.

[33] Christos P. Panteliadis. *Cerebral Palsy: A Multidisciplinary Approach*. Springer International Publishing, 3 edition, 2018.

[34] ES Draper, ID Gallimore, JJ Kurinczuk, PW Smith, T Boby, LK Smith, and BN Manktelow. *MBRRACE-UK – Perinatal Mortality Surveillance Report 2017*. The Infant Mortality and Morbidity Studies, 2018.

[35] Helen MacTier, Sarah Elizabeth Bates, Tracey Johnston, Caroline Lee-Davey, Neil Marlow, Kate Mulley, Lucy K. Smith, Meekai To, and Dominic Wilkinson. Perinatal management of extreme preterm birth before 27 weeks of gestation: A framework for practice. *Arch. of Disease in Childhood: Fetal and Neonatal Edition*, 105(3):F232–F239, 2020.

[36] Fatima Yousif Ismail, Ali Fatemi, and Michael V. Johnston. Cerebral plasticity: Windows of opportunity in the developing brain. *European Journal of Paediatric Neurology*, 21(1):23–48, 2017. Advances in Neuromodulation in Children.

[37] Iona Novak, Cathy Morgan, Lars Adde, James Blackman, Roslyn N. Boyd, Janice Brunstrom-Hernandez, Giovanni Cioni, Diane Damiano, Johanna Darrah, Ann-Christin Eliasson, Linda S. de Vries, Christa Einspieler, Michael Fahey, Darcy Fehlings, Donna M. Ferriero, Linda Fetters, Simona Fiori, Hans Forssberg, Andrew M. Gordon, Susan Greaves, Andrea Guzzetta, Mijna Hadders-Algra, Regina Harbourne, Angelina Kakooza-Mwesige, Petra Karlsson, Lena Krumlinde-Sundholm, Beatrice Latal, Alison Loughran-Fowlds, Nathalie Maitre, Sarah McIntyre, Garey Noritz, Lindsay Pennington, Domenico M. Romeo,

Roberta Shepherd, Alicia J. Spittle, Marelle Thornton, Jane Valentine, Karen Walker, Robert White, and Nadia Badawi. Early, Accurate Diagnosis and Early Intervention in Cerebral Palsy: Advances in Diagnosis and Treatment. *JAMA Pediatrics*, 171(9):897–907, 09 2017.

[38] Allison Shevell and Michael Shevell. Doing the "talk": Disclosure of a diagnosis of cerebral palsy. *Child Neurology*, 28(2):230–235, 2013.

[39] Dragan Zlatanović, Hristina Colovic, Zivkovic Vesna, Mirjana Kocic, Anita Stanković, Jelena Vučić, Nikola Bojovic, Maja Raičević, Mladjan Golubovic, Ljubomir Dinic, and Tamara Stanković. The importance of the prechtl method for ultra-early prediction of neurological abnormalities in newborns and infants. *AMM*, pages 111–115, 09 2019.

[40] Christa Einspieler, Heinz F.R. Prechtl, Fabrizio Ferrari, Giovanni Cioni, and Arend F. Bos. The qualitative assessment of general movements in preterm, term and young infants. *Early Human Development*, 50(1):47 – 60, 1997. Spontaneous Motor Activity as a Diagnostic Tool Functional Assessment of the Young Nervous System.

[41] Joan L. Lacey, Sian Rudge, Ingrid Rieger, and David A. Osborn. Assessment of neurological status in preterm infants in neonatal intensive care and prediction of cerebral palsy. *Australian Journal of Physiotherapy*, 50(3):137–144, 2004.

[42] Lilly Dubowitz, Eugenio Mercuri, and Victor Dubowitz. An optimality score for the neurologic examination of the term newborn. *The Journal of Pediatrics*, 133(3):406–416, 1998.

[43] Lena Krumlinde-Sundholm, Linda Ek, Elisa Sicola, Lena Sjöstrand, Andrea Guzzetta, Giuseppina Sgandurra, Giovanni Cioni, and Ann Christin Eliasson. Development of the Hand Assessment for Infants: evidence of internal scale validity. *Developmental Medicine and Child Neurology*, 59(12):1276–1283, 2017.

[44] Carol H. Leonard, Robert E. Piecuch, and Bruce A. Cooper. Use of the Bayley Infant Neurodevelopmental Screener with low birth weight infants. *Journal of Pediatric Psychology*, 26(1):33–40, 2001.

[45] Penelope B. Butler, Sandy Saavedra, Madeline Sofranac, Sarah E. Jarvis, and Marjorie H.

Woollacott. Refinement, reliability, and validity of the segmental assessment of trunk control. *Pediatric Physical Therapy*, 22(3):246–257, 2010.

[46] Christa Einspieler and Heinz F. R. Prechtl. Prechtl's assessment of general movements: A diagnostic tool for the functional assessment of the young nervous system, feb 2005.

[47] Carolina Yuri Panvequio Aizawa, Christa Einspieler, Fernanda Françoso Genovesi, Silvia Maria Ibidi, and Renata Hydee Hasue. The general movement checklist: A guide to the assessment of general movements during preterm and term age. *Jornal de Pediatria*, 2020.

[48] Christa Einspieler, Arend F. Bos, Melissa E. Libertus, and Peter B. Marschik. The general movement assessment helps us to identify preterm infants at risk for cognitive dysfunction. *Frontiers in Psychology*, 7, Mar 2016.

[49] Heinz FR Prechtl, Christa Einspieler, Giovanni Cioni, Arend F. Bos, Fabizi Ferrari, and Dieter Sontheimer. An early marker for neurological deficits after perinatal brain lesions. *The Lancet*, 349(9062):1361–1363, May 1997.

[50] Colleen Peyton and Christa Einspieler. General Movements: A Behavioral Biomarker of Later Motor and Cognitive Dysfunction in NICU Graduates. *Pediatric Annals*, 47(4):e159–e164, 2018.

[51] Eileen Ricci, Christa Einspieler, and Alexa K. Craig. Feasibility of Using the General Movements Assessment of Infants in the United States. *Physical and Occupational Therapy in Pediatrics*, 2018.

[52] Christa Einspieler and Heinz F. R. Prechtl. *Prechtls method on the qualitative assessment of general movements in preterm, term, and young infants*. Mac Keith Press, 2004.

[53] Cerebral Palsy Alliance. What is the general movements assessment?

[54] Amanda K.L. Kwong, Tara L. Fitzgerald, Lex W. Doyle, Jeanie L.Y. Cheong, and Alicia J. Spittle. Predictive validity of spontaneous early infant movement for later cerebral palsy: a systematic review. *Developmental Medicine and Child Neurology*, 60(5):480–489, 2018.

[55] Catherine Morgan, Cathryn Crowle, Traci-Anne Goyen, Caroline Hardman, Michelle Jackman, Iona Novak, and Nadia Badawi. Sensitivity and specificity of general movements assessment for diagnostic accuracy of detecting cerebral palsy early in an australian context. *Journal of Paediatrics and Child Health*, 52(1):54–59, 2016.

[56] Joan L. Lacey and David J. Henderson-Smart. Assessment of preterm infants in the intensive-care unit to predict cerebral palsy and motor outcome at 6 years. *Developmental Medicine and Child Neurology*, 40(5):310–318, 1998.

[57] Jade Kant. A comparison of the sensitivity and specificity of 3 neurological assessments currently in use on neonatal units. In *Developmental Medicine and Child Neurology*, 2013.

[58] Lilly Dubowitz and Victor Dubowitz. *The neurological assessment of the preterm and full-term newborn infant*. Mackeith Press, 2000.

[59] Domenico M. Romeo, Daniela Ricci, Claudia Brogna, and Eugenio Mercuri. Use of the Hammersmith Infant Neurological Examination in infants with cerebral palsy: A critical review of the literature, 2016.

[60] L Krumlinde-Sundholm, E Sicola, L Ek, A Guzetta, L Sjöstrand, G Cioni, and A Eliasson. The Hand Assessment for Infants, a new test for measuring use of hands and possible asymmetry in infants 3-10 months of age. AACPDM abstract. *Developmental Medicine and Child Neurology*, 54(I8):54–55, oct 2015.

[61] Kathleen H. Armstrong and Heather C. Agazzi. Chapter 2 - the bayley-iii cognitive scale. In Lawrence G. Weiss, Thomas Oakland, and Glen P. Aylward, editors, *Bayley-III Clinical Use and Interpretation*, Practical Resources for the Mental Health Professional, pages 29–45. Academic Press, San Diego, 2010.

[62] Sylvie Naar-King, Deborah A. Ellis, and Maureen A. Frey. *Assessing children's well-being: A handbook of measures*. Routledge, 2003.

[63] Michelle M. Macias, Conway F. Saylor, Margaret K. Greer, Jane M. Charles, Nancy Bell, and Lakshmi D. Katikaneni. Infant Screening: The Usefulness of the Bayley Infant Neurodevelopmental Screener and the Clinical Adaptive Test/Clinical Linguistic Auditory

Milestone Scale. *Journal of Developmental and Behavioral Pediatrics*, 19(3):155–161, 1998.

[64] Ryan Cunningham, María B. Sánchez, Penelope B. Butler, Matthew J. Southgate, and Ian D. Loram. Fully automated image-based estimation of postural point-features in children with cerebral palsy using deep learning. *Royal Society Open Science*, 6(11):191011, 2019.

[65] Lisbeth Hansen, Katrine Thingholm Erhardsen, Jesper Bencke, Stig Peter Magnusson, and Derek John Curtis. The reliability of the segmental assessment of trunk control (satco) in children with cerebral palsy. *Physical & Occupational Therapy In Pediatrics*, 38(3):291–304, 2018. PMID: 28749721.

[66] Veronka Horber, Ute Grasshoff, Elodie Sellier, Catherine Arnaud, Ingeborg Krägeloh-Mann, and Kate Himmelmann. The role of neuroimaging and genetic analysis in the diagnosis of children with cerebral palsy. *Frontiers in Neurology*, 11, 2021.

[67] Ingeborg Krageloh-Mann and Veronka Horber. The role of magnetic resonance imaging in elucidating the pathogenesis of cerebral palsy: a systematic review. *Developmental Medicine & Child Neurology*, 49(2):144–151, 2007.

[68] Anna Kovalchuk and Bryan Kolb. Low dose radiation effects on the brain–from mechanisms and behavioral outcomes to mitigation strategies. *Cell Cycle*, 16(13):1266–1270, 2017.

[69] Summer Kaplan and Ammie M. White. *Cranial Ultrasound in Cerebral Palsy*, pages 101–111. Springer International Publishing, Cham, 2018.

[70] Dimitrios I. Zafeiriou, Eleftherios E. Kontopoulos, and Ioannis Tsikoulas. Characteristics and prognosis of epilepsy in children with cerebral palsy. *Journal of Child Neurology*, 14(5):289–294, 1999. PMID: 10342595.

[71] Elsa Tillberg, Bengt Isberg, and Jonas K.E. Persson. Hemiplegic (unilateral) cerebral palsy in northern Stockholm: Clinical assessment, brain imaging, EEG, epilepsy and aetiologic background factors. *BMC Pediatrics*, 20(1), 2020.

[72] Mintaze Kerem Gunel. Physiotherapy for Children with Cerebral Palsy. *Epilepsy in Children - Clinical and Social Aspects*, 2011.

[73] Thomas Michael O'Shea. Diagnosis, treatment, and prevention of cerebral palsy. *Clinical Obstetrics and Gynecology*, 51(4):816–828, dec 2008.

[74] Hobbs Rehabilitation Neurological Specialists. Cerebral palsy: Hobbs neurological rehabilitation, 2021.

[75] Toril Fjørtoft, Christa Einspieler, Lars Adde, and Liv Inger Strand. Inter-observer reliability of the "Assessment of Motor Repertoire - 3 to 5 Months" based on video recordings of infants. *Early Human Development*, 85(5):297–302, 2009.

[76] Royal Society of Great Britain. *Machine learning : the power and promise of computers that learn by example*, volume 66. The Royal Society, 2017.

[77] Issam El Naqa and Martin J. Murphy. *What Is Machine Learning?*, pages 3–11. Springer International Publishing, Cham, 2015.

[78] Christopher-John Farrell. Identifying mislabelled samples: Machine learning models exceed human performance. *Annals of Clinical Biochemistry*, 58(6):650–652, 2021. PMID: 34210147.

[79] Ohad Oren, Bernard J. Gersh, and Deepak L. Bhatt. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *The Lancet Digital Health*, 2(9):e486–e488, 2020.

[80] Pablo Duboue. *The Art of Feature Engineering: Essentials for Machine Learning*. Cambridge University Press, 2020.

[81] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning, Second Edition*. The MIT Press, US, Cambridge, 2018.

[82] Michael W Berry, Azlinah Mohamed, and Bee Wah Yap. *Supervised and unsupervised learning for data science*. Springer, 2019.

[83] David Fumo. Types of machine learning algorithms you should know, Jun 2017.

[84] Ankur A Patel. *Hands-on unsupervised learning using Python: how to build applied machine learning solutions from unlabeled data*. O'Reilly Media, 2019.

[85] Julianna Delua. Supervised vs. unsupervised learning: What's the difference?, Mar 2021.

[86] IBM. Machine learning, July 2020.

[87] IBM. Ai vs. machine learning vs. deep learning vs. neural networks: What's the difference?, May 2020.

[88] Amitha Mathew, P. Amudha, and S. Sivakumari. Deep learning techniques: An overview. In Aboul Ella Hassanien, Roheet Bhatnagar, and Ashraf Darwish, editors, *Advanced Machine Learning Technologies and Applications*, pages 599–608, Singapore, 2021. Springer Singapore.

[89] Vanessa Buhrmester, David Münch, and Michael Arens. Analysis of explainers of black box deep neural networks for computer vision: A survey. *CoRR*, abs/1911.12116, 2019.

[90] Sambit Mahapatra. Why Deep Learning over Traditional Machine Learning? — by Sambit Mahapatra — Towards Data Science, 2018.

[91] Hyeoun Ae Park. An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, 43(2):154–164, 2013.

[92] William S. Noble. What is a support vector machine? *Nature Biotechnology*, 24(12):1565–1567, 2006.

[93] Paul E. Utgoff. Incremental Induction of Decision Trees. *Machine Learning*, 4(2):161–186, 1989.

[94] T Gaber, A Tharwat, A Ibrahim, and AE Hassanien. Linear discriminant analysis: A detailed tutorial. *USIR*, pages 0–22, 2017.

[95] Yoav Freund and Robert E. Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[96] Jerome Friedman, Robert Tibshirani, and Trevor Hastie. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2):337–407, 2000.

[97] Evelyn Fix and J. L. Hodges. Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review / Revue Internationale de Statistique*, 57(3):238–247, 1989.

[98] Robert Kozma, Cesare Alippi, Yoonsuck Choe, Francesco, and Carlo Morabito. *Artificial Intelligence in the Age of Neural Networks and Brain Computing*. Academic Press, 11 2018.

[99] Joanne S Katz. Interrater and Intrarater Reliability Using Prechtl's Method of Qualitative Assessment of General Movements in Infants. *International Journal of Pediatric Research*, 2(1):13–16, 2016.

[100] I. Bernhardt, M. Marbacher, R. Hilfiker, and L. Radlinger. Inter- and intra-observer agreement of prechtl's method on the qualitative assessment of general movements in preterm, term and young infants. *Early Human Development*, 87(9):633–639, 2011.

[101] Franziska Heinze, Katharina Hesels, Nico Breitbach-Faller, Thomas Schmitz-Rode, and Catherine Disselhorst-Klug. Movement analysis by accelerometry of newborns and infants for the early detection of movement disorders due to infantile cerebral palsy. *Medical and Biological Engineering and Computing*, 48(8):765–772, 2010.

[102] Mohan Singh and Donald J. Patterson. Involuntary gesture recognition for predicting cerebral palsy in high-risk infants. In *International Symposium on Wearable Computers (ISWC) 2010*, pages 1–8, 2010.

[103] Dana Gravem, M. Singh, C. Chen, Julia Rich, J. Vaughan, Ken Goldberg, F. Waffarn, Pai H. Chou, D. Cooper, D. Reinkensmeyer, and Donald Patterson. Assessment of infant movement with a compact wireless accelerometer system. *Journal of Medical Devices-transactions of The Asme*, 6:021013, 2012.

[104] Mingming Fan, Dana Gravem, Dan M. Cooper, and Donald J. Patterson. Augmenting gesture recognition with erlang-cox models to identify neurological disorders in premature

babies. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, page 411–420, New York, NY, USA, 2012. Association for Computing Machinery.

[105] Dominik Karch, Keun-Sun Kang, Katarzyna Wochner, Heike Philippi, Mijna Hadders-Algra, Joachim Pietz, and Hartmut Dickhaus. Kinematic assessment of stereotypy in spontaneous movements in infants. *Gait & Posture*, 36(2):307–311, 2012.

[106] Heike Philippi, Dominik Karch, Keun-Sun Kang, Katarzyna Wochner, Joachim Pietz, Hartmut Dickhaus, and Mijna Hadders-Algra. Computer-based analysis of general movements reveals stereotypies predicting cerebral palsy. *Developmental Medicine and Child Neurology*, 56(10):960–967, 2014.

[107] Hodjat Rahmati, O. Aamo, Ø. Stavdahl, R. Dragon, and L. Adde. Video-based early cerebral palsy prediction using motion segmentation. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3779–3783, 2014.

[108] Mikkel D. Olsen, Anna Herskind, Jens Bo Nielsen, and Rasmus R. Paulsen. Using motion tracking to detect spontaneous movements in infants. In Rasmus R. Paulsen and Kim S. Pedersen, editors, *Image Analysis*, pages 410–417, Cham, 2015. Springer International Publishing.

[109] Mikkel Damgaard Olsen, Anna Herskind, Jens Bo Nielsen, and Rasmus Reinhold Paulsen. Model-based motion tracking of infants. In Lourdes Agapito, Michael M. Bronstein, and Carsten Rother, editors, *Computer Vision - ECCV 2014 Workshops*, pages 673–685. Springer International Publishing, 2015.

[110] H. Rahmati, H. Martens, O. M. Aamo, O. Stavdahl, R. Stoen, and L. Adde. Frequency analysis and feature reduction method for prediction of cerebral palsy in young infants. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(11):1225–1234, Nov 2016.

[111] Archana Machireddy, Jan Santen, Jenny Wilson, Julianne Myers, Mijna Hadders-Algra, and Xubo Song. A video/imu hybrid system for movement estimation in infants. In *EMBC 2017*, volume 2017, pages 730–733, 07 2017.

[112] L. Meinecke, N. Breitbach-Faller, C. Bartz, R. Damen, G. Rau, and C. Disselhorst-Klug. Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy. *Human Movement Science*, 25(2):125–144, 2006.

[113] Nao Kanemaru, Hama Watanabe, Hideki Kihara, Hisako Nakano, Tomohiko Nakamura, Junji Nakano, Gentaro Taga, and Yukuo Konishi. Jerky spontaneous movements at term age in preterm infants who later developed cerebral palsy. *Early Human Development*, 90(8):387–392, 2014.

[114] Lars Adde, Jorunn Helbostad, Alexander R. Jensenius, Mette Langaas, and Ragnhild Støen. Identification of fidgety movements and prediction of CP by the use of computer-based video analysis is more accurate when based on two video recordings. *Physiotherapy Theory and Practice*, 29(6):469–475, 2013.

[115] Lars Adde, Hong Yang, Rannei Sæther, Alexander Refsum Jensenius, Espen Ihlen, Jia yan Cao, and Ragnhild Støen. Characteristics of general movements in preterm infants assessed by computer-based video analysis. *Physiotherapy Theory and Practice*, 34(4):286–292, 2018.

[116] Annette Stahl, Christian Schellewald, Øyvind Stavdahl, Ole Morten Aamo, Lars Adde, and Harald Kirkerod. An optical flow-based method to predict infantile cerebral palsy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(4):605–614, 2012.

[117] Silvia Orlandi, Kamini Raghuram, Corinna R. Smith, David Mansueto, Paige Church, Vibhuti Shah, Maureen Luther, and Tom Chau. Detection of Atypical and Typical Infant Movements using Computer-based Video Analysis. *in EMBC*, 2018:3598–3601, 2018.

[118] Kamini Raghuram, Silvia Orlandi, Vibhuti Shah, Tom Chau, Maureen Luther, Rudaina Banihani, and Paige Church. Automated movement analysis to predict motor impairment in preterm infants: a retrospective study. *Journal of Perinatology*, 39(10):1362–1369, 2019.

[119] Espen Ihlen, Ragnhild Støen, Lynn Boswell, Raye-Ann Deregnier, Toril Fjørtoft, Deborah Gaebler-Spira, Catherine Labori, Marianne Loennecken, Michael Msall, Unn Möinichen, Colleen Peyton, Michael Schreiber, Inger Silberg, Nils Thomas Songstad, Randi Vågen, Gunn Øberg, and Lars Adde. Machine learning of infant spontaneous movements for the

early prediction of cerebral palsy: A multi-site cohort study. *Journal of Clinical Medicine*, 9, 12 2019.

[120] Walter Baccinelli, Maria Bulgheroni, Valentina Simonetti, Francesca Fulceri, Angela Caruso, Letizia Gila, and Maria Luisa Scattoni. Movidea: A software package for automatic video analysis of movements in infants at risk for neurodevelopmental disorders. *Brain Sciences*, 10(4), 2020.

[121] William Schmidt, Matthew Regan, Micheal Fahey, and Andrew Paplinski. General movement assessment by machine learning: why is it so difficult? *Journal of Medical Artificial Intelligence*, 2(0), 2019.

[122] Toshio Tsuji, Shota Nakashima, Hideaki Hayashi, Zu Soh, Akira Furui, Taro Shibanoki, Keisuke Shima, and Koji Shimatani. Markerless Measurement and Evaluation of General Movements in Infants. *Scientific Reports*, 10(1):1–13, 2020.

[123] Masahiko Suzuki, Hiroshi Mitoma, and Mitsuru Yoneyama. Quantitative Analysis of Motor Status in Parkinson's Disease Using Wearable Devices: From Methodological Considerations to Problems in Clinical Applications. *Parkinson's Disease*, 2017(1), 2017.

[124] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *CVPR 2011*, pages 1297–1304, June 2011.

[125] Liuyang Zhou, Zhiguang Liu, Howard Leung, and Hubert P. H. Shum. Posture reconstruction using kinect with a probabilistic model. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, VRST '14, pages 117–125, New York, NY, USA, Nov 2014. ACM.

[126] Pierre Plantard, Antoine Muller, Charles Pontonnier, Georges Dumont, Hubert P. H. Shum, and Franck Multon. Inverse dynamics based on occlusion-resistant kinect data: Is it usable for ergonomics? *International Journal of Industrial Ergonomics*, 61:71–80, 2017.

[127] Steffi L Colyer, Murray Evans, Darren P Cosker, and Aki I T Salo. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. *Sports Med - Open*, 2018.

[128] Lars Adde, Jorunn L. Helbostad, Alexander Refsum Jensenius, Gunnar Taraldsen, and Ragnhild Støen. Using computer-based video analysis in the study of fidgety movements. *Early Human Development*, 85(9):541–547, 2009.

[129] Lars Adde, Jorunn L. Helbostad, Alexander R. Jensenius, Gunnar Taraldsen, Kristine H. Grunewaldt, and Ragnhild StØen. Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study. *Developmental Medicine and Child Neurology*, 52(8):773–778, 2010.

[130] Kamini Raghuram, Silvia Orlandi, Paige Church, Tom Chau, Elizabeth Uleryk, Petros Pechlivanoglou, and Vibhuti Shah. Automated movement recognition to predict motor impairment in high-risk infants: a systematic review of diagnostic test accuracy and meta-analysis. *Developmental Medicine and Child Neurology*, 63(6):637–648, 2021.

[131] Joonsoo Lee and Al Bovik. Chapter 19 - video surveillance. In Al Bovik, editor, *The Essential Guide to Video Processing*, pages 619–651. Academic Press, Boston, 2009.

[132] Thomas Brox, Christoph Bregler, and Jitendra Malik. Large displacement optical flow. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 41–48, 2009.

[133] Pavan Turaga, Rama Chellappa, and Ashok Veeraraghavan. Advances in video-based human activity analysis: Challenges and approaches. In Marvin V. Zelkowitz, editor, *Advances in Computers*, volume 80 of *Advances in Computers*, pages 237–290. Elsevier, 2010.

[134] Salwa O. Slim, Ayman Atia, Marwa M.A. Elfattah, and Mostafa-Sami M.Mostafa. Survey on human activity recognition based on acceleration data. *International Journal of Advanced Computer Science and Applications*, 10(3), 2019.

[135] Nida Saddaf Khan and Muhammad Sayeed Ghani. *A Survey of Deep Learning Based Models for Human Activity Recognition*, volume 120. Springer US, 2021.

[136] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 40(04):834–848, apr 2018.

[137] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339, 2014.

[138] Jian Dong, Qiang Chen, Xiaohui Shen, Jianchao Yang, and Shuicheng Yan. Towards unified human parsing and pose estimation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 843–850, 2014.

[139] Ke Gong, Xiaodan Liang, Dongyu Zhang, Xiaohui Shen, and Liang Lin. Look into person: Self-supervised structure-sensitive learning and a new benchmark for human parsing, 2017.

[140] Fangting Xia, Peng Wang, Xianjie Chen, and Alan L. Yuille. Joint multi-person pose estimation and semantic part segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[141] George Papandreou, Tyler Zhu, Liang-Chieh Chen, Spyros Gidaris, Jonathan Tompson, and Kevin Murphy. Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model, 2018.

[142] Hao-Shu Fang, Guansong Lu, Xiaolin Fang, Jianwen Xie, Yu-Wing Tai, and Cewu Lu. Weakly and semi supervised human body part parsing via pose-guided knowledge transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 70–78, 06 2018.

[143] Amy Bearman, Olga Russakovsky, Vittorio Ferrari, and Li Fei-Fei. What's the point: Semantic segmentation with point supervision. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 549–565, Cham, 2016. Springer International Publishing.

[144] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1, June 2005.

[145] S. Lo and A. Tsoi. Human action recognition: A dense trajectory and similarity constrained latent support vector machine approach. In *2013 2nd IAPR Asian Conference on Pattern Recognition*, pages 230–235, Nov 2013.

[146] H. Wang, A. Kläser, C. Schmid, and C. Liu. Action recognition by dense trajectories. In *CVPR 2011*, pages 3169–3176, June 2011.

[147] Jingtian Zhang, Hubert P. H. Shum, Jungong Han, and Ling Shao. Action recognition from arbitrary views using transferable dictionary learning. *IEEE Transactions on Image Processing*, 27(10):4709–4723, 2018.

[148] L. Xia, C. Chen, and J. K. Aggarwal. View invariant human action recognition using histograms of 3d joints. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 20–27, June 2012.

[149] Andreas Holzinger, Chris Biemann, Constantinos S. Pattichis, and Douglas B. Kell. What do we need to build explainable ai systems for the medical domain?, 2017.

[150] Nikolas Hesse, Christoph Bodensteiner, Michael Arens, Ulrich G. Hofmann, Raphael Weinberger, and A. Sebastian Schroeder. Computer vision for medical infant motion analysis: State of the art and RGB-D data set. In *Computer Vision - ECCV 2018 Workshops*. Springer International Publishing, 2018.

[151] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2019.

[152] Hiroshi Akima. A new method of interpolation and smooth curve fitting based on local procedures. *J. ACM*, 17(4):589–602, October 1970.

[153] Q. Wu, G. Xu, F. Wei, L. Chen, and S. Zhang. Rgb-d videos-based early prediction of infant cerebral palsy via general movements complexity. *IEEE Access*, 9:42314–42324, 2021.

[154] Manli Zhu, Qianhui Men, Edmond S. L. Ho, Howard Leung, and Hubert P. H. Shum. Interpreting deep learning based cerebral palsy prediction with channel attention. In *IEEE International Conference on Biomedical and Health Informatics (BHI'21)*. IEEE, 2021.

[155] Binh Nguyen-Thai, Vuong Le, Catherine Morgan, Nadia Badawi, Truyen Tran, and Svetha Venkatesh. A spatio-temporal attention-based model for infant movement assessment from videos. *IEEE Journal of Biomedical and Health Informatics*, 25(10):3911–3920, 2021.

[156] Dimitrios Sakkos, Kevin D. Mccay, Claire Marcroft, Nicholas D. Embleton, Samiran Chattopadhyay, and Edmond S. L. Ho. Identification of abnormal movements in infants: A deep neural network for body part-based prediction of cerebral palsy. *IEEE Access*, 2021.

[157] Muhammad Tausif Irshad, Muhammad Adeel Nisar, Philip Gouverneur, Marion Rapp, and Marcin Grzegorzek. AI approaches towards prechtl's assessment of general movements: A systematic literature review. *Sensors (Switzerland)*, 20(18):1–32, 2020.

[158] W. Rueangsirarak, J. Zhang, N. Aslam, E. S. L. Ho, and H. P. H. Shum. Automatic musculoskeletal and neurological disorder diagnosis with relative joint displacement from human gait. *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, 26(12):2387–2396, Dec 2018.

[159] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. Learning actionlet ensemble for 3d human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5):914–927, 2014.

[160] Davide Chicco and Giuseppe Jurman. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1):1–13, 2020.

[161] Nikolas Hesse, Sergi Pujades, Javier Romero, Michael J. Black, Christoph Bodensteiner, Michael Arens, Ulrich G. Hofmann, Uta Tacke, Mijna Hadders-Algra, Raphael Weinberger, Wolfgang Müller-Felber, and A. Sebastian Schroeder. Learning an infant body model from RGB-D data for accurate full body motion analysis. In *MICCAI*. Springer, 2018.

[162] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms, 2012.

[163] Anezka Kazikova, Michal Pluhacek, and Roman Senkerik. How does the number of objective function evaluations impact our understanding of metaheuristics behavior? *IEEE Access*, 9:44032–44048, 2021.

[164] Leena Haataja, Eugenio Mercuri, Rivka Regev, Frances Cowan, Mary Rutherford, Victor Dubowitz, and Lilly Dubowitz. Optimality score for the neurologic examination of the infant at 12 and 18 months of age. *The Journal of Pediatrics*, 135(2):153–161, 1999.

[165] K. D. McCay, E. S. L. Ho, C. Marcroft, and N. D. Embleton. Establishing pose based features using histograms for the detection of abnormal infant movements. In *EMBC 2019*, pages 5469–5472, July 2019.

[166] K. D. McCay, E. S. L. Ho, H. P. H. Shum, G. Fehringer, C. Marcroft, and N. D. Embleton. Abnormal infant movements classification with deep learning on pose-based features. *IEEE Access*, 2020.

[167] Wei Wei, Edmond Ho, Kevin McCay, Robertas Damaševičius, Rytis Maskeliūnas, and Anna Esposito. Assessing facial symmetry and attractiveness using augmented reality. *Pattern Analysis and Applications*, 2021.

[168] Jingtian Zhang, Hubert P. H. Shum, Kevin McCay, and Edmond S. L. Ho. Prior-less 3d human shape reconstruction with an earth mover's distance informed cnn. In *Motion, Interaction and Games*, MIG '19, New York, NY, USA, 2019. Association for Computing Machinery.