Évaluation précoce du développement des enfants prématurés par analyse quantitative des mouvements généraux et estimation de la pose 3D par apprentissage profond

Ameur Soualmi

N°d'ordre NNT : 2023STET058

THÈSE de DOCTORAT
DE L'UNIVERSITÉ JEAN MONNET SAINT-ÉTIENNE

Membre de l'Université de LYON

École Doctorale N° 488
Sciences, Ingénierie, Santé (SIS)

Spécialité : **Informatique**

Soutenue publiquement le 15/12/2023, par :
Ameur SOUALMI

EARLY DEVELOPMENTAL EVALUATION OF PREMATURE INFANTS WITH QUANTITATIVE GENERAL MOVEMENTS ASSESSMENT USING DEEP LEARNING BASED 3D POSE ESTIMATION

Devant le jury composé de :

DUCOTTET, Christophe / PR / LHC- UMR CNRS 5516, Université Jean Monnet, Saint-Etienne	Président
CARRAULT, Guy PR / LTSI - Inserm U 1099, Université de Rennes I	Rapporteur
GOUIFFES, Michèle HDR/ LISN UMR CNRS 9015, Polytech Paris Saclay	Rapporteuse
BEURTON-AIMAR, Marie HDR / LABRI UMR CNRS 5800, Université de Bordeaux	Examinatrice
JOUEN, François HDR / R2P2, USVQ, AP-HP, EPHE-PSL, Ecole Pratique des Hautes Etudes	Examinateur

PATURAL, Hugues PU-PH / UMR 1059-SAINBIOSE-DVH, Université Jean Monnet, Saint-Etienne	Directeur de thèse
ALATA, Olivier PR / LHC- UMR CNRS 5516, Université Jean Monnet, Saint-Etienne	Co-directeur de thèse
GIRAUD, Antoine MCU-PH / UMR 1059-SAINBIOSE-DVH, Université Jean Monnet, Saint-Etienne	Encadrant

Happiness can exist only in acceptance.
George Orwell

Dedication

To my beloved mom ...

Aknowledgement

I would like to express my gratitude to my supervisors for their valuable support and mentorship throughout the course of this thesis. Their dedication to finding the necessary resources and their unwavering support during challenging times played an integral role in the successful completion of this thesis. I am deeply appreciative of Antoine, who was more than a mentor and helped me a lot during the past three years. Additionally, I extend my sincere thanks to the Neonatal unit staff, who welcomed me into their team and showed great patience with me. I would also like to extend my gratitude to my life mentor, Dr. Cherifi Dalila, for her continuous support and guidance. Finally, I want to express my thanks to my brother, Bilal, for his support from the day I started my master's degree, and to my wonderful girlfriend, who has been a constant source of motivation.

Droits d'auteurs

ABSTRACT

The General Movement assessment (GMA) is a validated assessment of premature infants' brain maturation primarily based on the qualitative analysis of the complexity, variability, and fluidity of infants' spontaneous motor activity and their classification as normal or abnormal movements. The GMA can identify preterm infants presenting an early abnormal developmental trajectory before term-equivalent age, which permits a personalized early developmental intervention. However, GMA is time-consuming and relies on a subjective analysis; these limitations restrict the implementation of GMA in clinical practice. Automatic GMA has been explored in the last two decades, thanks to the continuous progress in computer vision methods. However, it remains limited due to the lack of training data and outdated automatic movement detection methods. In this thesis, we introduce the AGMA study (IDRCB 2020-A03335-34), based on a dataset of 282 videos from 129 preterm infants, born before 33 weeks of gestational age (GA) and hospitalized in the Neonatology department of the Centre Hospitalier Universitaire de Saint-Etienne, France. Over 88k images from this dataset were used to fine-tune state-of-the-art deep learning models for 2D infant pose estimation task. Then, based on the new stereoscopic 3D pose estimation framework presented in this study, we achieved a mean error of 1.72 cm in the 3D infant pose estimation on 18000 stereoscopic images. These deep-learning models were used to create a software that automatically selects general movements sequences from long video recordings. Furthermore, the research introduced a novel quantitative parameter called Mean 3D Dispersion (M3D), based on a new spherical representation method, able to identify infants displaying abnormal movements with 73% accuracy. This thesis addresses the present-day challenges for automatically assessing infants' general movements and makes a step further toward early predicting infants' neurodevelopmental outcome.

RÉSUMÉ

L'analyse de l'activité motrice spontanée de l'enfant prématuré selon la méthode de Prechtl, appelé évaluation des mouvements généraux, est une évaluation validée de la maturation cérébrale de l'enfant. Elle est basée sur l'analyse qualitative de la complexité, la variabilité, et la fluidité, de l'activité motrice spontanée, ainsi que sa classification comme normale ou anormale. Cette méthode peut identifier de manière précoce les nourrissons prématurés présentant une trajectoire de développement anormale, ce qui permet une intervention précoce et personnalisée. Cependant, cette méthode est chronophage et repose sur une analyse subjective ; ces limites restreignent sa mise en œuvre dans la pratique clinique. L'analyse automatisée de la motricité spontanée a été explorée au cours des deux dernières décennies, grâce aux progrès continus dans les méthodes de traitement d'image et d'intelligence artificielle. Cependant, elle reste limitée en raison du manque de données d'entraînement et de méthodes obsolètes de détection automatique des mouvements. Dans cette thèse, nous présentons l'étude AGMA (IDRCB 2020-A03335-34), basée sur une base de données qui contient 282 vidéos de 129 nourrissons prématurés, nés avant 33 semaines d'aménorrhée, hospitalisés dans le service de néonatalogie du Centre Hospitalier Universitaire de Saint-Étienne, en France. Plus de 88 000 images de cet ensemble de données ont été utilisées pour réentraîner les plus récents réseaux d'apprentissage profond afin d'estimer la posture 2D des nourrissons de manière automatique. Ensuite, en utilisant le nouveau système stéréoscopique d'estimation de la posture 3D présenté dans cette étude, nous avons obtenu une erreur moyenne de 1,72 cm dans l'estimation de la posture 3D sur 18 000 images stéréoscopiques de nourrissons. Ces réseaux d'apprentissage profond ont ensuite été utilisés pour créer un logiciel de sélection automatique des séquences d'analyse de la motricité spontanée des enfants prématurés. De plus, nous avons introduit un nouveau paramètre quantitatif appelé 3D Mean Dispersion (M3D) capable de discriminer les mouvements normaux et anormaux des nourrissons. Ce paramètre est basé sur une nouvelle représentation sphérique. Il permet d'identifier les nourrissons présentant des mouvements anormaux avec une précision de 73%. Cette thèse répond aux défis actuels de l'évaluation automatique des mouvements généraux des enfants prématurés et constitue une avancée significative dans la conception d'un outil pronostique précoce du développement de l'enfant prématuré.

Content

Listing of Figures

Listing of Tables

Abbreviations

AAP – American Academy of Pediatrics.
API – Application Programming Interface.
AUC – Area Under the Curve.

CA – Corrected Age.
CI – Confidence Interval.
CNN – Convolutional Neural Netwrok.
CP – Cerebral Palsy.
CS – Cramped Synchronized.
CUS – Cranial Ultrasound.

DBN – Dynamic Bayesian Network.
DT – Decision Trees.

EEG – Electroencephalogram.
EMTS – Electromagnetic Tracking Systems.

FN – False Negatives.
FP – False Positives.
FPS – Frame Per Second.

GA – Gestational Age.
GMA – General movements assessment.
GMs – General Movements.
GUI – Graphical User Interface.

HOJD2D – Histogram of Joint Displacement 2D.
HOJO2D – Histogram of Joint Orientation 2D.

LDA – Linear Discriminant Analysis.

LDOF – Large Displacement Optical Flow.
LLGMN – Linearized Gaussian Mixture Network.
LOVO – Leave-One-Video-Out.
LSTM – Long Short-Term Memory.

mAP – mean Average Precision.
MINI-RGBD – Moving INfants In RGB-D.
MRI – Magnetic Resonance Imaging.

NB – Naive Bayes.

OF – Optical Flow.
OKS – Object Keypoint Similarity.

PAF – Part Association Field.
PCK – Percentage of Correct Key-points.

RF – Random Forest.
ROC – Receiver Operating Characteristic.

SMIL – Skinned Multi-Infant Linear Model.
SOTA – state-of-the-art.
SVM – Support Vector Machine.

TN – True Negatives.
TNR – True Negative Rate.
TP – True Positives.
TPR – True Positive Rate.

WHO – World Health Organization.

1

Introduction

Premature birth, defined as birth occurring before 37 weeks of gestation, is associated with a higher risk of death and developmental disorders such as cerebral palsy, cognitive deficits, and sensory impairments. Regarding these serious risks, the identification of early signs of developmental impairments is crucial for early individual intervention, which is associated with better developmental outcomes [1]. The Prechtls' method of General movements assessment (GMA) [2] was proved to be a valuable tool with a high degree of reliability in identifying potential neurodevelopmental issues in premature infants. This method is based on the qualitative analysis of the infant's spontaneous movements that can be visualized from birth until the appearance of voluntary goal-directed movements at approximately 4 months post-term. However, this method is subjective and time-consuming. In the last two decades, many studies searched for GMA automation as a solution to clinical GMA limitations, especially with the recent advances in computer vision and deep learning techniques. The third chapter reviews these studies and presents the different computer vision techniques used in the last two decades to detect infants' movements. These methods can be sensor-based using motion and vision sensors attached to infants' bodies, and sensor-free using only RGB and depth cameras. Deep learning methods became a trend due to their performance, particularly in human pose estimation (HPE) tasks. Some research teams [3]–[6] used existing models to directly estimate infants' poses from images, but they were inaccurate since they were trained on datasets containing only images of adults. Others [7]–[9] created datasets to retrain some models, but the images were collected from the internet or synthetic which are not representative of preterm infants. So, these methods are still limited and face multiple challenges. Therefore, we made a general summary of all these methods with the various challenges in this field of study at the end of the third chapter in addition to the goals of this study aiming to improve the existing methods to automatically assess infants' general movements.

This thesis introduces a new approach to automatically assess infants' spontaneous movements. It consists of a new stereoscopic framework to estimate the infants' poses

in 3D. The different steps followed are detailed in the fourth chapter including the video acquisition protocol used in this study and the AGMAPose dataset collected and manually annotated to fine-tune state-of-the-art deep learning models for infants' 2D pose estimation. The same chapter compares different retrained deep learning networks to existing solutions and shows the advantage of the presented framework over other 3D pose estimation methods such as depth cameras. This framework allowed a complete 3D analysis of infants' movements in space. Therefore, chapter 5 introduces a new software that was created to automatically select general movements sequences from long videos, which reduces the time needed to perform GMA in clinical practice. Moreover, a new quantitative parameter called mean 3D dispersion (M3D) was introduced. This automatic classification parameter was achieved based on a new spherical representation of joints' movements in space that allowed a normalized and objective comparison of infants' movements with different weights and ages. The performance of this parameter in discriminating Normal and Abnormal movements according to GMA was validated on the AGMA dataset. This dataset was collected over almost 3 years, containing 282 video recordings of 129 premature infants born <33 weeks Gestational Age (GA). All the details about this dataset and its labeling are given in the same chapter. The thesis ends with a conclusion that summarizes the important findings of this study, in addition to the future perspectives and directions to be explored in order to better automatically assess infants' general movements.

2

Premature Birth and General Movements Assessment

Outline

2.1 Introduction

The goal of this chapter is to review the developmental aspect of premature birth and the relevance of the infant's GMA. This chapter explores the risks associated with preterm birth, including motor impairments such as cerebral palsy, cognitive deficits, and sensory impairments like visual and hearing issues. In the second part, this chapter presents Prechtl's method of GMA. This assessment method provides important insights into the neurological development of infants through the analysis of their spontaneous movements. This method categorizes general movements as normal or abnormal based on their characteristics, offering a mean to identify potential neurodevelopmental issues in infants. The chapter presents the procedure and reliability of GMA, focusing on its significance in early identification and intervention for infants at risk in addition to its limitations as a clinical tool.

2.2 Premature Birth

2.2.1 Definition

The World Health Organization (WHO) defines premature birth, also known as preterm birth, as birth before 37 completed weeks of gestation instead of full-term birth at approximately 40 weeks. Premature birth is defined as extreme (less than 28 weeks), very preterm (between 28 and 31 weeks), and mild preterm (between 32 and 36 weeks) [10]. In 2020, and according to the WHO, 10% of newborns were born premature, which represents approximately 13.4 million babies worldwide and concerns 55 000

babies per year in France [11]. Premature birth is a source of concern because it is associated with a higher risk of death or serious disability (see Figure 2.1). It is the second leading cause of death among children under 5 years old, as well as a common cause of infant death in the first 12 hours after birth [12].

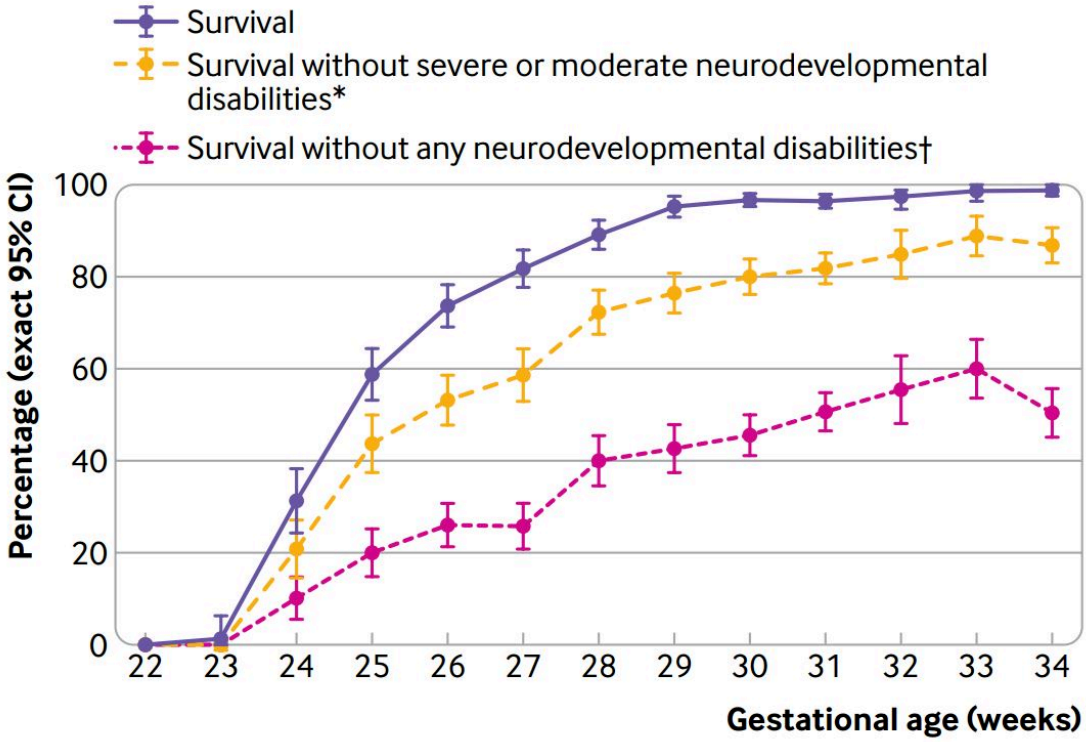


Figure 2.1: Survival rates among preterm born children according to the EPIPAGE-2 study [13].

2.2.2 Causes

Recent studies have highlighted many factors thought to contribute to preterm birth. Even if all the reasons that some babies are born prematurely are not known and understood, several risk factors can increase the risk that a woman will have a preterm delivery. Depending on the causes, the premature birth can be either spontaneous or indicated.

Spontaneous Premature birth

Spontaneous preterm birth occurs when labor begins on its own before the full-term period. This can be induced by several factors. The main factor of spontaneous preterm

birth is the inflammation or infection of the placenta, called chorioamnionitis [1]. Even if individual behaviors have not been proven to have a strong influence on the risk of preterm birth, a healthy lifestyle is associated with favorable pregnancy outcomes [14], [15]. There is a reduced risk of preterm birth among women who have a good diet and do not smoke tobacco [16]. Alcohol consumption is also an important factor in premature birth; women who have more than one drink per day have an increased risk of preterm birth [17]. Drug use was also studied as a preterm birth factor, but women who are positive for drug use may be at higher risk of preterm birth based on the factors that marked them as potential drug users rather than based on the use of drugs per se [18]. High levels of stress during pregnancy can increase the risk of premature birth. In a sample of 5,873 Danish women, Hedegaard et al. found that experiencing one or more highly stressful life events between 16 and 30 weeks of pregnancy was associated with a higher risk of preterm delivery [19]. Also, Life events, such as divorce, a death in the family, illness, injury, or the loss of a job, were associated with preterm birth [18].

Indicated Premature birth

Indicated preterm birth is when a doctor determines that it is medically necessary to deliver the baby before full-term due to specific maternal or fetal health concerns. A number of maternal and fetal medical conditions are associated with an increased risk of preterm birth. Chronic medical conditions such as diabetes and preeclampsia may lead to an indicated delivery [20]. Also, fetal heart rate abnormalities and growth retardation are common indications for preterm delivery.

2.2.3 Developmental Outcomes

Preterm birth is a major contributor to infant mortality [21]. According to the study from Pierrat et al. [13], the survival rate of premature infants in France before 22 weeks is close to 0, while at 23 weeks it is around 1%, at 24 weeks around 31%, and at 25 weeks about 59%. Thanks to improvements in obstetric and neonatal

intensive care in the last decades, there has been an important decline in infant and neonatal mortality. However, surviving preterm infants usually have disabilities because premature birth interrupts in-utero development; all of the organs can be present, but they are still immature. Therefore, preterm infants are born with an immature brain [22]. Even if the formation of brain structures occurs in the early months of pregnancy, brain development, maturation, and the establishment of all neural connections primarily occur during the third trimester. The children can have serious long-term neurodevelopmental impairments, including Cerebral Palsy (CP), cognitive impairments, visual and hearing impairments, and behavioral disorders [18].

Cerebral Palsy

CP appears in the early years of life due to a non-progressive disorder of the central nervous system, resulting in a group of permanent disorders of the development of movement and posture, causing activity limitation [23]. The severity of CP is determined by its type, which limbs are affected, and the degree of functional limitations ranging from mild to severe CP [24]. Also, CP percentage among survivors is inversely proportional with gestational age: 0.1% to 0.2% of full-term children develop CP, whereas 11% to 12% born at 27 to 32 weeks of gestation and 7% to 17% born at less than 27 or 28 weeks of gestation develop CP [25], [26]. Spastic diplegia is a common type of CP in premature infants; it involves an involuntary increase of muscle tone (tightness or stiffness) that primarily affects the limbs, with legs and feet being much more affected than the arms and hands [27]. Other types of CP can also occur, such as spastic hemiplegia, in which movements of the limbs are on one side of the body but not on the other, in addition to spastic quadriplegia where the impairment concerns all four limbs, with the head and trunk involved as well. CP diagnosis may not become certain until the second year of life [13]

Cognitive Impairment

The brain's gray matter, which contains the cell bodies of neurons responsible for cognitive functions, is particularly affected by preterm birth and the incomplete development of the brain. Premature infants often have less developed cerebral cortex, which plays a critical role in functions like memory, attention, language, and problem-solving [18]. Thus, it results in cognitive impairments that can have lasting effects on children's development and quality of life. Intellectual developmental disorders include difficulties with problem-solving, reasoning, and learning, which can significantly impact their academic progress [28]. Also, they can suffer from behavioral issues such as impulsivity and hyperactivity and struggle with social interactions, finding it challenging to initiate or maintain relationships with others [18].

Visual and hearing impairments

The visual and hearing systems pass through a critical development in the womb during the last part of pregnancy. So premature birth can cause serious visual impairments like retinopathy of prematurity (ROP), a condition in premature infants affecting the retina, the light-sensitive tissue at the back of the eye responsible for vision. Strabismus, also known as crossed eyes or misalignment of the eyes is a condition in premature infants. In addition, they can suffer from hearing loss due to auditory nerve damage. Marlow et al. study on 6-year-old children born in 1995 at gestational ages of less than 26 weeks stated that 3% had a profound sensorineural hearing impairment that could not be corrected with hearing aides, 3% had a sensorineural hearing impairment that could be corrected with hearing aids, and 4% had a mild hearing impairment [29].

2.2.4 Developmental Evaluation

Regarding the risks of premature birth, the identification of early signs of developmental issues is crucial. This early detection allows early interventions and targeted individual

monitoring [1]. Neuroimaging methods such as Cranial Ultrasound (CUS) and Magnetic Resonance Imaging (MRI) are used to identify preterm infants with significant brain injuries and central nervous system (CNS) abnormalities. The American Academy of Pediatrics (AAP) recommends that preterm neonates born before 30 weeks GA should be screened with an early CUS examination on days 7–10 [30]. In addition to these neuroimaging methods, a functional evaluation of the premature infant's brain activity can be made using an Electroencephalogram (EEG). Abnormal EEG patterns in premature infants, especially those associated with seizures or continuous abnormal patterns, are associated with neurodevelopmental disorders [31], such as CP or cognitive impairments. Another functional brain evaluation of preterm infants is the Prechtls' method of GMA. This non-invasive method is based on the analysis of the infant's spontaneous movements. It categorizes general movements as normal or abnormal based on their characteristics, offering a mean to identify early neurodevelopmental trajectories of infants. In their review (see Table 2.1), Bosanquet et al. [32] concluded that GMA is the most predictive method, with the best sensitivity 98% (95% CI 74–100%) and specificity 91% (95% CI 83–93%) for predicting CP in the fidgety period, 3 months post-term age. Also, a higher predictive validity for CP has been demonstrated with GMA than with CUS and MRI [32], [33].

2.3 The General Movements Assessment Method

2.3.1 Background

Dr. Heinz Prechtl, an Austrian pediatrician and neuroscientist, was the leading researcher on the GMA method in the early 70s. He accomplished extensive research and clinical work on the spontaneous movement patterns in preterm and full-term infants. He suggested that these spontaneous movements had a prenatal history [34], and later confirmed it with the advancement of ultrasound technology. These movements

are not triggered by sensory input [35], are considered a significant expression of the developing brain, and form the cornerstone of early development [36].

2.3.2 Infants' General Movements

General Movements (GMs) are defined as a series of gross movements of variable speed and amplitude, which involve all body parts but lack a distinctive sequencing of the participating body parts [37]. GMs have a gradual onset and end, wax and wane in force, intensity, and speed. Rotations along the axis of the limbs and slight changes in the direction of movements make them fluent and elegant. This creates the impression of complexity and variability. GM patterns can vary from a few seconds to a couple of minutes [37] and can evolve over time and slowly be replaced by voluntary goal-directed movements at approximately 4 months post-term. These movements are characterized by their complexity, variability, and fluidity, and their normality is considered as a positive sign of typical development. Prechtl's method of GMA classifies the movements as normal or abnormal according to their quality and also their chronological appearance (see Figure 2.2) [2].

Normal GMs

Normal GMs before term are called fetal or preterm GMs, and they may appear different from those of full-term infants due to the unique developmental context of premature birth. Preterm GMs may be of large amplitude and fast speed [38], while conserving the complex, variable, and fluid character. From term age (40 weeks post-menstrual) until four weeks post-term age, they are called writhing movements. They are characterized by small to moderate amplitudes and slow to moderate speeds. Their elliptical presentation creates the impression of a writhing quality [2]. When writhing movements start disappearing at around 9 weeks post-term age, fidgety movements become more apparent. Fidgety movements are small, jerky, and irregular movements that involve the entire body; they can include quick flexion and extension of limbs, head jerks, and

torso. These movements are present and continual in the awake infant and absent during fussing or crying. Fidgety movements can be observed until 20 weeks post-term, when intentional and antigravity movements begin to dominate [39].

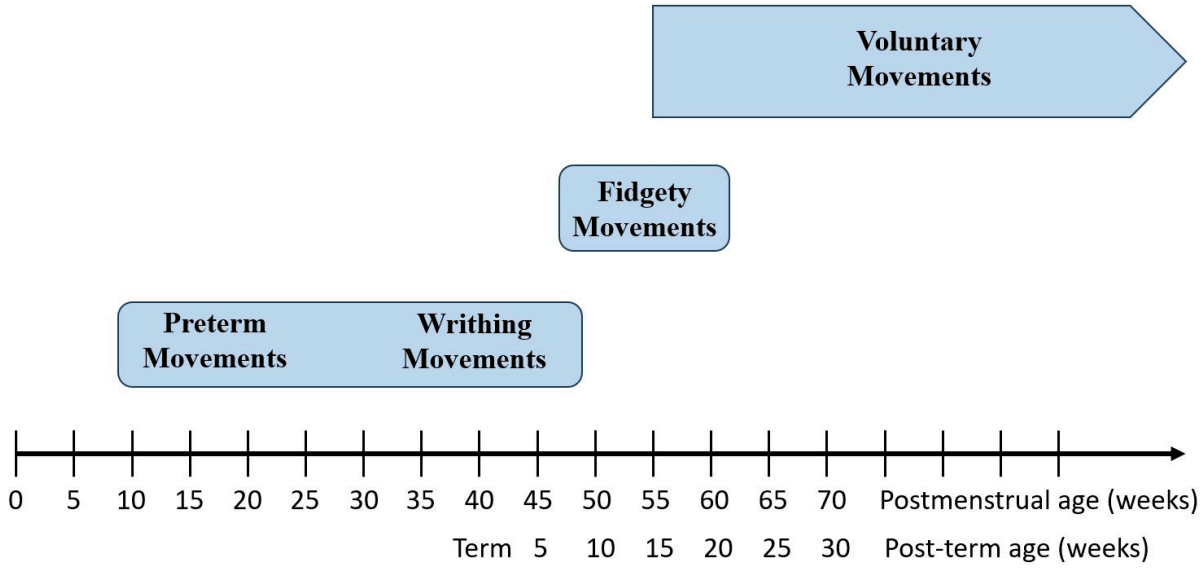


Figure 2.2: Developmental course of general movements

Abnormal Gms

If the nervous system is impaired, GMs lose their complex, variable, or fluid character. This defines poor repertoire or are cramped-synchronized or chaotic GMs. This holds true for the preterm, term, and early post-term age (first two months), whereas fidgety movements can be either absent or abnormal [2]. Unlike normal GMs, poor repertoire GMs are repetitive, and the movements in the different body parts do not occur in a complex manner [2]. They lack variability and are characterized by a limited set of movements. This form of abnormal GMs can be observed during preterm, term, and early post-term age. Cramped-synchronized GMs are associated with certain neurological conditions like CP. A distinct pattern of movements is present, where the infant's limbs, especially the arms and legs, appear to be tightly flexed or cramped in a synchronized manner. Movements appear rigid and lack the normal, smooth, and fluent character [39], [40]. The chaotic GMs are characterized by movements that lack

the typical smoothness, coordination, and fluency seen in normal infant movements. In a chaotic manner, all the limbs are continuously jerky and of large amplitudes. However, their appearance is rare. Fidgety movements can be absent or abnormal. Abnormal fidgety movements look like fidgety movements, but their amplitude, speed, and jerkiness are greatly exaggerated. Abnormal fidgety movements are rarely observed and have very low predictive value [39], [41]. Absent fidgety movements, as defined by their name, are associated with the absence of fidgety movements from 9 to 20 weeks post-term. The absence of fidgety movements is highly predictive of later neurological impairments, such as CP. Cramped-synchronized movements start to develop at 34 weeks post-menstrual age (PMA). If cramped-synchronized movements are present at three to four months post-term age, fidgety movements will be absent [39].

2.3.3 Procedure

As Einspieler at al.[2] suggested, infants GMs are better observed by video recordings even if it is simpler to assess the motor activity by directly observing the movements with the unaided eye. Video recordings can be played back and stored for future observation and reference. The team also described the ideal video acquisition setup: infants should be lying in a supine position in the incubator, bed, or on the floor. If the assessment is done without acoustic signals, the assessor must be able to see the infant's face to make sure that rigid movements are not due to crying. The room temperature should be comfortable, and in case of prolonged episodes of fussing, crying, or distraction, the recording must be stopped since it interferes with GMs' quality. In infants older than 36 weeks, recordings should preferably be performed during state 4 [42], which is characterized by open eyes, irregular respiration, present movements, and absent crying. Younger preterm infants should be recorded when bouts of activity occur, irrespective of whether the infant is awake or sleeping [42]. The duration of the recording depends on the age of the infant. To collect about three GMs for a reliable assessment, the video recording varies from 30 to 60 minutes. Later, this recording is

viewed at a fast speed, and about three GM sequences are extracted for assessment. The assessment depends on visual Gestalt perception; the first step is to classify the GMs as normal or abnormal. In case of abnormality, they are classified as poor repertoire, chaotic, cramped-synchronized, abnormal fidgety movements, and absence of fidgety movements [2]. An individual developmental trajectory can be obtained if GMs are assessed multiple times. These trajectories will indicate the consistency or inconsistency of normal or abnormal findings [37].

2.3.4 Reliability

The strong validity and reliability of the GMs are demonstrated in many studies [33], [41], [42]. The inter-observer reliability has been demonstrated by several groups at 90% agreement [43]. The review from Noble et al. [44] reported that the assessment of GMs had a sensitivity (see section 5.4.4) from 83.3% to 100% and specificity (see section 5.4.4) between 59.1% and 100% in predicting high-risk CP and learning disability, depending on whether term or preterm infants were being assessed and at what age. Bosanquet et al. [32] concluded that GMA is the most predictive, with the best sensitivity 98% (95% Confidence Interval (CI) 74–100%) and specificity 91% (95% CI 83–93%) of predicting CP in the fidgety period, at 3 months post-term age (see Table 2.1). Also, a higher predictive validity for CP has been demonstrated with GMs than with cranial ultrasound (CUS) and magnetic resonance imaging (MRI) [32], [33]. Other studies have shown that GMA is more predictive of infants' neurological outcomes when it is done as a series of multiple GMs assessments over multiple weeks. In one early study, it was found that infants with persistently abnormal movements assessments (at writhing and fidgety GM phases) had an increased likelihood (70-85%) of developing CP [34].

2.3.5 Limitations

The GMA is a valuable tool for assessing the neurological and developmental status of infants. However, like any assessment method, it has its limitations and considerations

Test	Number of reviewed studies	Number of participants	Sensitivity % (95% CI)	Specificity % (95% CI)
Neurological assessment	4	1190	88	87
CUS	10	2827	74	92
MRI	3	1190	86 to 100	89 to 97
GMA	6	1358	98	91

Table 2.1: Predictive value of neurodevelopmental assessments used to predict cerebral palsy in young children [32]. CI: Confidence interval.

primarily due to the need for skilled clinicians. These clinicians must undergo extensive training and accumulate years of practical assessment experience to achieve a certain level of accuracy. This is one of the reasons why GMA has not yet been established widely in clinical routines. In addition, as explained in the GMA procedure, GMA depends on gestalt visual perception of movements [2], which is a subjective interpretation of an observer who assesses the quality and characteristics of a baby's general movements. This can result in variability between observers in how to interpret and assess the same set of movements. Moreover, the GMA procedure is a very time-consuming task. From a long video recording, clinicians must manually extract about three GM sequences [2]. This task includes visualizing all the videos and trying to manually select 1 to 3 minutes video sequences where the infants are moving the most. This time-consuming process is an obstacle for clinicians to practice GMA in clinical routines.

2.4 Conclusion

In this chapter, we have observed the significant risks that premature birth imposes on preterm infants' development, ranging from motor impairments to cognitive deficits and sensory issues. Among the assessment methods discussed, Prechtl's method of GMA appears as a valuable tool with a high degree of reliability, offering the potential for early identification of neurodevelopmental impairments. However, this assessment

method has many limitations, including the need for experienced assessors, subjectivity in interpretation, and its time-consuming nature. Regarding these challenges, more studies focused on the task of GMA automation, especially with the recent advances in computer vision and deep learning techniques.

3

Review of Automatic General Movements Assessment Methods

Outline

3.1 Introduction

To automatically assess infants' movements, a crucial initial step involves the automatic quantification and tracking of these movements. In the early stages of research in this domain, various sensors such as accelerometers, electromagnetic tracking systems, and reflective markers were employed. However, due to the limitations associated with these methods, researchers developed sensor-free methods based on different computer vision algorithms like optical flow. Over the past decade, with the advancements in deep learning algorithms, particularly in human pose estimation applications, many studies have opted for convolutional neural networks to track infants' movements automatically. While the majority of studies focused on 2D methodologies, others explored techniques for comprehensive 3D analysis. On the other hand, once the infants' movements were detected and tracked, studies tried to classify these movements based on multiple techniques, including machine and deep learning methods, with varying results and accuracies.

The goal of this chapter is twofold. First, to review all the methods proposed over the last two decades. As a conclusion of this review, the second goal is to expose the main objectives of this thesis work.

3.2 Sensor-Based Methods

3.2.1 Motions Sensors

Motion sensors were widely used for human gesture recognition and human-computer interaction applications. Therefore, it inspired many other applications in the medical field, such as detecting characteristic motions of Parkinson's disease [45], detecting stimming movements in autistic children [46], in addition to automatic general movements assessment. Gravem et at.[47] used 5 accelerometers to assess the movements of 10 preterm infants with a mean gestational age of 36.3 weeks GA. Accelerometers were embedded in cloth bands that were placed around the wrists, ankles, and forehead of the infants. They measured three orthogonal axes of acceleration on the head and each of the four limbs. They performed feature extraction, resulting in feature vectors containing 166 statistical and temporal features. These features were used to identify Cramped-synchronized movements based on Support Vector Machine (SVM), Decision Trees (DT), and Dynamic Bayesian Network (DBN). They reported an accuracy between 70% and 90%, but the small number of subjects in this study hardly supports the results. Singh et al. [48] used a similar approach with 4 accelerometers (see Figure 3.1). They enrolled 10 infants between 30–43 weeks GA of age. Their goal was to identify Cramped Synchronized (CS) movements using Naive Bayes, Support Vector Machines, and a pruned Decision Tree. They classified the movements as normal vs. abnormal based only on the presence and absence of CS movements, which is theoretically inaccurate. Heinze et al. [49] included 23 infants (19 healthy with a mean gestational age of 39.6 weeks) and four high-risk (mean gestational age of 29.25 weeks). A classifier based on a decision tree algorithm was implemented to differentiate between healthy and pathological data. They reported an overall accuracy of 89%. However, their dataset was too unbalanced, and the 4 accelerometers used were wired ones. In addition to accelerometers as a motion sensor, Electromagnetic Tracking Systems (EMTS) were also employed. Karch et al. [50] attached 8 sensors to the right side of the body for

Figure 3.1: Baby in the NICU with accelerometers on each limb [48].

20 infants between term and 3 months post-term age. They used simple movements of the hand and the foot to calculate the joint centers and the joint axes of a multi-segmental model. They evaluated the recording accuracy by considering the differences between recorded and estimated sensor positions and orientations. Philippi et al. [51] used eight Electromagnetic Tracking Systems placed in two configurations. First, the sensors were attached to one side of the body on the upper and the lower limbs. For the second configuration, the sensors were attached to both upper limbs. This was performed at the corrected age of 3 months for 49 preterm at high-risk infants and 18 term-infants at low risk. They used a statistical analysis with the Area Under the Curve (AUC) to compare the general predictive performance of features to predict CP. Despite their accuracy (relatively with small datasets), the weight and size of motion and wearable sensors are significant obstacles in infant movement tracking and are time-consuming as an approach. Therefore, more recent studies have focused on developing methods based on visual sensors.

3.2.2 Visual Sensors

Visual sensors replaced motion sensors in the GMA context since they do not affect the movement of the infant by their weight or their size. Meinecke et al. [52] used reflecting markers with a Vicon 370 motion analysis system equipped with 6 cameras (see Figure

3.2 and 3.3.b). They included twenty-two infants, 15 healthy with a mean birth age of 39.7 weeks GA, and seven high-risk preterm with a mean birth age of 28.6 weeks GA. The measurements were acquired at the fourth week Corrected Age (CA) and allowed to extract 53 parameters describing the quality of movements. However, the group has tried to find the most significant 8 parameters from this selection using cluster analysis with Euclidean distance. Then, they used the quadratic discriminant analysis algorithm to classify movements as normal or abnormal based on the 8 parameters. Even if the group reported 73% accuracy, their setup is very difficult to implement in a clinical environment since it needs a special acquisition protocol and a specific location to be chosen for the reflective markers at each recording. Kanemaru et al. [53] used a similar approach with reflective markers easier to attach to the infant's body than the previous study (see Figure 3.3.c) but limited to 2D analysis. They analyzed the spontaneous movements at 36–44 weeks postmenstrual age (PMA) for 145 preterm infants (born 22–36 weeks PMA). Sixteen of the infants developed CP by 3 years of age, while 129 developed normally. They calculated multiple parameters, including the average velocity of limb movements, number of movement units, kurtosis of acceleration, etc. Their goal was to verify whether the movements of infants who were later diagnosed with CP were jerky compared to those of healthy ones, so they did a simple statistical comparison between the parameters for each population (CP vs. Healthy). Berthouze et al. [46] have used a total of 37 reflective markers, with a frame on each joint that includes 3 or 4 markers (see Figure 3.3.a). This was to avoid self-occlusion when using 6 cameras. Regarding the multiple challenges opposed by the methods based on visual sensors, including their cost and feasibility, researchers focused more on totally sensor-free methods.

3.3 Sensor-Free AGMA

To automatically asses infants' general movements without motion sensors, many studies presented new methods based on computer vision tools. Adde et al [56]

Figure 3.2: Infant motion capture system [54].

Figure 3.3: Reflective markers for movements detection. (a) Reflective markers type 1 [55]. (b) Reflective markers type 2 [54]. (c) Reflective markers type 3 [53].

included 82 in their study and video-recorded their GMs between 10 and 18 weeks post-term age (Fidgety movements period). The GMs were classified following Prechtl's method of GMA as present, abnormal, or absent. They used motion images, which were calculated as the difference between two consecutive frames where each pixel represents a point value of 0 and 1, 0 being black and representing no movement, and 1 being white and representing movement [56]. These motion images were used to calculate Motiongrams as the average of the motion images to a 1D vector over time,

Motion quantities as the motion image pixel mean, and motion centroids as the spatial center of the positive pixels in the motion image. These parameters were used for statistical analysis and comparison between the two populations (present vs. absent fidgety movements), such as Kolmogorov–Smirnov test, t-tests, and logistic regression. Even if the study was the first to automatically classify infants' movements according to the GMA, the frame differencing method is very sensitive to noise and does not deliver information about the directions and the speeds of the infants' movements. A step further was made with the use of Optical Flow (OF) algorithms. This method calculates the motion vectors of points or features between consecutive video frames, providing information about how objects move and change position. Stahl et al. [57] enrolled 82 infants aged 10–18 weeks (Fidgety movements period) post-term, and 15 of them were confirmed diagnosed with CP after two years of age. Infants were recorded multiple times in a total of 136 recordings. They performed motion extraction using optical flow describing the displacement vector field of pixels between consecutive image frames. After that, they opted for feature extraction using wavelets and frequency analysis, then proceeded to the classification of the movements using SVM to discriminate between unimpaired infants and infants with CP. Many other studies opted for Large Displacement Optical Flow (LDOF) [58]–[60]. Orlandi et al.[58] collected 127 videos of 98 infants with normal motor repertoire and 29 infants with abnormal movements, according to GMA. The videos were recorded between 3-5 months of corrected age (Fidgety movements period). They started with a motion estimation with LDOF, then applied a skin model mask to remove the background and segment infants' bodies. After that, they extracted 643 numerical features from each video and performed feature selection to reduce the number of features. These features were used in different classifiers, Logistic, AdaBoost, LogitBoost, and Random Forest (RF), to classify GMs as typical and atypical, and diagnose CP (CP vs. not-CP). Raghuram et al. [59] continued the previous study with the same subjects and features extraction method, including LDOF, segmentation, etc. However, they aimed to predict motor impairment rather

than just CP. They used only 3 parameters for MI prediction: the minimum velocity, the mean velocity of the infant's silhouette, and the mean vertical velocity. Even if these traditional computer vision methods (OF, LDOF, etc.) are more cost and time-effective than sensor-based methods, they still have many limitations, especially in the case of occlusions and illumination changes.

Schmidt et al.[61] were the first to use a neural network to classify the fidgety movements in 7 classes (Normal, Intermittent, not classifiable, Needs follow-up, Sporadic, Absent, and Abnormal) from a dataset of 500 videos. They collected 2,445 video segments of 5-20 seconds duration and sampled them at 5 frames intervals to reduce the data size. They fine-tuned the first 8 layers of the Keras VGG19 model, pre-trained on ImageNet database and passed the features through a max pooling layer, normalized them, and to an Long Short-Term Memory (LSTM) layer for classification of the image sequences (145 images per video). The team has reported 65.1% accuracy using a 10-fold cross-validation, a sensitivity of 50.8%, and specificity of 27.4%. Commenting on these poor results, the group reported that such results were not unexpected as their project was in the early stages, and with further work, a more robust result will be expected.

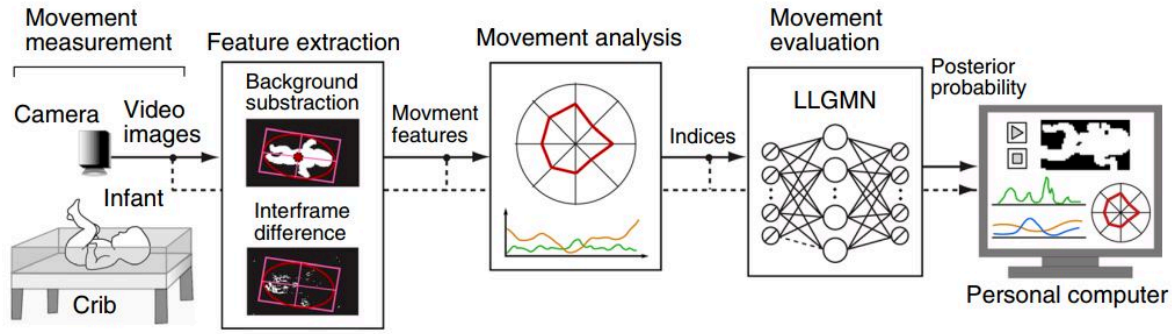


Figure 3.4: GMs classification framework according to Tsuji et al.[62].

Tsuji et al.[62] introduced another method to classify the movements of 21 subjects, consisting of 3 steps (see Figure 3.4). They start with feature extraction using frame differencing and segmentation to calculate four main movement characteristics: movement magnitude, movement balance, movement rhythm, and movement of the body

center. These features were fed to a Linearized Gaussian Mixture Network (LLGMN), which estimates the probabilistic distribution of every data point. They classified the movements into 4 classes (Writhing movements, Fidgety movements, Poor Repertoire, and CS) with an accuracy of 83.1%. The manually extracted features in this work do not give a lot of information about the directions and the speeds of the infants' movements. Tsuji et al. suggested the inclusion of additional details focusing on the pose estimation methods in the directions and future work.

3.4 2D Pose Estimation

Human pose estimation is the precise detection of anatomical landmarks in the human body. This task plays an important role in equipping machines with a comprehensive method of human actions and is now a fundamental challenge within computer vision and its affiliated domains.

3.4.1 Human 2D Pose Estimation

In the last decade, deep learning has revolutionized the task of pose estimation for adults and achieved very promising results on different state-of-the-art (SOTA) datasets (e.g., COCO [63] or MPII [64]). Deep learning human pose estimation methods can be categorized into two main categories: Bottom-Up and Top-Down methods.

Bottom-Up Method

This method first estimates the locations of all the body joints in the image and then groups them together to form human poses. The first step is typically performed using a heatmap-based method, which predicts a heatmap for each body joint, and the second step is performed using a grouping algorithm, which groups together the body joints that belong to the same person (e.g.[65]–[69]). Many methods rely on statistical and geometric models to estimate occluded joints [70], [71], associative methods such as Part Association Field (PAF) [65], or associative embedding [72]. Another approach is using a library of known poses, known as anchor poses [73], but this could limit the generalization ability of the model and its capacity to detect unseen poses. One major challenge in bottom-up methods is associating detected keypoints with the person present in the image. Toward this goal, Iqbal and Gall [74] built a convolutional pose machine-based pose estimator to estimate the joint candidates. Then, they used integer linear programming to solve the joint-to-person association problem and obtain human body poses even in severe occlusions. Openpose[65] is considered a breakthrough

approach for grouping keypoints with the introduction of PAF. Although this method has achieved impressive results on high-resolution images, it has poor performance with low-resolution images and occlusions. Kreisset et al. [68] addressed this problem and proposed a bottom-up method called PifPaf that uses a Part Intensity Field to predict the locations of body parts and a Part Association Field to represent the joints association. In more recent work, Cheng et al. [75] combined the HRNet structure with multi-resolution pyramids to obtain multi-scale features in the HigherHRNet network to tackle the problem of scale variation in bottom-up pose estimation. However, one of the limitations of bottom-up methods is the use of global contextual information of the image, which can increase the runtime complexity.

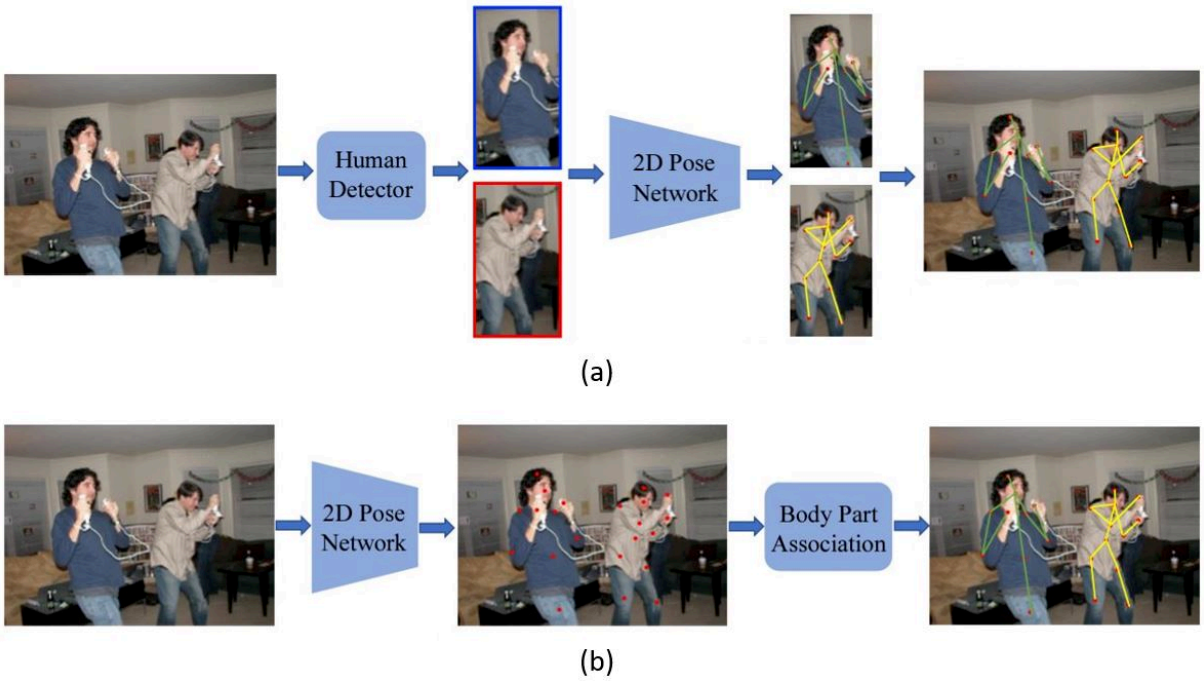


Figure 3.5: 2D human pose estimation methods. (a)Top-down approach, (b)Bottom-up approach. [76]

Top-Down Method

The top-down methods [75], [77]–[81] for human pose estimation consist of two steps: the process starts by detecting human bodies in the image and then estimating the

poses of the detected bodies (see Figure 3.5). The first step is mainly performed using an object detector, and the second is performed using a pose estimator. The latter, particularly the heatmap-based pose estimators, can also be classified according to their network architecture. Classical pose estimators use an encoder-decoder design network, where an encoder takes an image as input and extracts features using a series of narrowing convolution blocks. As the data flows through the encoder, the feature maps produced by each layer become progressively more abstract and capture information at different spatial scales. Then, the decoder gradually upsamples the feature maps from the encoder and outputs heatmaps, indicating the likelihood of the keypoints' presence at various locations in the image. The stacked hourglass architecture [80] used many small encoder-decoder architectures as hourglass modules and stacked them on top of each other with skip connections. The skip connections allowed the stacked hourglass architecture to learn global and local features. Hourglass networks were further expanded to incorporate the multi-context approach in [82] by augmenting the backbone with residual units. However, this complex architecture increases the processing process, making it computationally expensive. Aiming to offer a multi-scale approach to feature representations, HRNet [81] includes both high and low resolutions to keep high-resolution representations throughout the network. Zhang et al.[83] introduced the DARK method with a new heatmap decoding process, which increased the accuracy of HRNet. The disadvantage of top-down human pose estimation methods is that the person detector may fail to identify the boundaries of largely overlapped human bodies in multi-person scenes.

3.4.2 Infant 2D Pose Estimation

The advances in human pose estimation inspired many works on infant pose estimation. One of the widely used deep learning models is OpenPose [65], which was trained on adult images only. However, it was widely used to estimate infants' poses in the automatic general movements context [3]–[6]. This can affect the accuracy of body

keypoints detection since the model was trained on datasets containing only images of adults who are structurally and anatomically different compared to infants [84]. Reich et al.[3] have included 51 newborns in their study, aiming to detect the presence of fidgety movements (present vs absent). Their method consists of estimating the infants' pose from videos using Openpose; the pose contains 25 keypoints with pixel coordinates (x,y). Then, they used these detections over 50 frames for each video to train an SMNN classification network (shallow multilayer neural network). Moccia et al.[85]introduced a fully convolutional neural network to estimate 2D infants' poses from 2D depth images. They used 5 layers and an output layer, giving 12 confidence maps for limb joints and 8 affinity fields for joint connections. Their training and testing sets had 1080 images each, which we find very small for training a CNN.

Doroniewicz et al. [6] video recorded 31 infants born full-term (38–42 weeks). A group of experts classified their GMs as normal writhing movements (17) and poor repertoire (14) based on GMA method. Their methods starts by pose estimation using openpose. Then they extract some kinematic features for movements trajectories such as Factor of Movements Area, Factor of Movements Shape, and Center of Movement's Area. Features vectors were used to train multiple binary classifiers, including SVM with the kernel, Random Forests (RF), and a Linear Discriminant Analysis (LDA). They reported similar accuracies with the 3 classifiers (around 80%), but we find that the amount of training data is very small. To address the inaccuracy of Openpose for infant pose estimation, Chambers et al. [7] have retrained the Openpose network with their own labeled dataset of infant pose consisting of 9039 infant images collected partially from videos on Youtube and clinical data, which reduced the mean error by 60%. In addition to the fact that Openpose lacks sufficient scaling of network depth and it is computationally insufficient [86], the ground truth body joints annotation process using Vatic [7] was not very accurate and was missing many occluded and not visible joints as shown in their publicly available dataset. The retrained model was used along with the YouTube and clinical dataset, which contained videos of full-term infants filmed between

3 and 11 months of age, to extract 38 kinematic features (posture, acceleration, velocity, etc.) and train a Naive Bayes (NB) classifier to detect unhealthy when their movements are different from the healthy reference dataset. The use of YouTube videos of infants with unknown backgrounds (age, status, etc.) is not reliable for medical diagnostics; in addition, their approach discriminates if infants are different from the healthy reference group but does not explain how they are different. The absence of open-source infants pose images mainly for privacy concerns is a problem that many researchers have tried to address. Hesse et al. [8] introduced a synthetic dataset of infants, called Moving INfants In RGB-D (MINI-RGBD), created from only 12 RGB-D videos of moving infants using a textural mapping to a 3D model mesh called Skinned Multi-Infant Linear Model (SMIL) model. Both its simplicity and exclusively synthetic nature cause the pose estimation models trained on MINI-RGBD not to generalize well on real-world infant images [9]. This is why Huang et al. [9] created a hybrid synthetic and real infant pose dataset (SyRIP) from real images collected from YouTube and Google images. They used a similar approach to SMIL by fitting the 3D model mesh to a real infant image and getting synthetic images with changing backgrounds, texture maps, lightning, and camera position, resulting in 700 real and 1000 infant synthetic images. This small dataset was used to train an architecture (FiDIP) that contains a domain classifier that promotes a feature extractor to retain the ability to extract keypoints information but also to ignore the differences between the real and the synthetic input images [9]. The group tested the effect of retraining three existing models as backbone networks on their SyRIP dataset in different ways, but what we found questionable was mainly the small number of images used for testing (100 & 500) and the use of mean Average Precision (mAP) evaluation metric with OKS thresholds [63] which was previously implemented specifically for the COCO dataset (explained in section 4.4.3). McCay et al.[87] made use of the MINI-RGBD dataset by examining the video sequences in this synthetic dataset and classifying them as normal and abnormal according to GMA. They used Openpose to estimate the infants' poses through the video sequences and extracted the

displacement of 14 keypoints. They created some kinematic features such as Histogram of Joint Orientation 2D (HOJO2D) and Histogram of Joint Displacement 2D (HOJD2D) to train a number of classifiers, including k-Nearest Neighbour, LDA, and the Ensemble classifier. Using leave-one-out cross-validation, an accuracy of 91.67% was achieved in the classification of infants' movements task. However, the dataset used here was fully synthetic and contained only 12 videos. In addition, there was no information about the quality of movements (preterm, or writhing, or. . .). The same team has continued this study with the same method [88], but instead of using machine learning classifiers, the feature vectors computed by HOJO2D, HOJD2D, were used to train a neural network and a Convolutional Neural Netwrok (CNN) architecture and compared their results with the machine learning algorithms used previously. They reported that deep learning methods are more performing than classical machine learning algorithms.

3.5 3D Pose Estimation

3.5.1 Human 3D Pose Estimation

3D human pose estimation offers a better understanding of human actions and movements than 2D pose estimation since it provides information in the three dimensions of space rather than two dimensions. Therefore, many studies addressed the challenge with different methods.

3D Pose Estimation From Monocular Images

Recovering a 3D human pose from a single image is appealing due to the low requirement of the image. Many researchers [89]–[93] suggested straightforward methods in which 3D pose is estimated by an end-to-end network without intermediately estimating 2D pose representations. Martinez et al.[94] were the first to do 2D poses lifting to 3D with a simple yet very effective neural network. With the progress of 2D human pose estimation algorithms, many studies [95]–[97] have proceeded with 2D pose estimation

in the first stage, then 2D to 3D lifting to get 3D pose in the second stage. However, the method could fail due to over-reliance on the 2D pose detector. Therefore, more recent studies focused on how to fuse 2D joint heatmaps with 3D image cues to reduce that ambiguity. Tekin et al.[98] used one stream of their network that accounted for the 2D joint locations and the corresponding uncertainties and a second stream that leveraged all 3D image cues by directly acting on the image. The outputs of these two streams were then fused to obtain the final 3D human pose estimate. However, estimating the 3D pose from a monocular image is challenging due to the partial occlusion problem. One way to overcome this problem is by estimating 3D human pose from multiple views since the occluded part in one view may become visible in other views.

3D Pose Estimation From Multi-View Images

Unlike the 3D pose estimation from monocular images method, which estimates pose from only one image, other studies [99]–[104] have opted for methods using multiple views images. This method can reduce the pose estimation error significantly. However, it is challenging to fuse information from multiple views (see Figure 3.6). Kocabas et al.[103] introduced a self-supervised learning method called Epipolar Pose for 3D human pose estimation. During training, Epipolar Pose estimates 2D poses from multi-view images and then utilizes epipolar geometry to obtain 3D poses, which are subsequently used to train a 3D pose estimator. Iskakov et al. [104] used another method with multi-view images. After showing the drawbacks of learnable algebraic triangulation, they proposed an approach based on volumetric triangulation, where they project the feature maps produced by multiple 2D pose estimation networks into 3D volumes. This was done by filling a 3D cube around the person via projecting output of the 2D network along projection rays inside the 3D cube. The cubes obtained from multiple views were then aggregated and processed by a 3D convolutional neural network to output 3D heatmaps. Even though more image views can lead to better 3D pose estimations, this can be computationally expensive and very sensitive since it relies on precise camera

calibration and synchronization. Any errors or discrepancies in these parameters can lead to inaccuracies in the 3D pose estimation. In the same year, Qiu et al. [105] fed multi-view images into a CNN model to merge information from other views into the current one. Furthermore, they proposed a recursive pictorial structure model to optimize the 3D poses, which can progressively reduce the quantization error to obtain better results. All the previous studies have shown that estimating 3D poses from multi-view images is more accurate than from monocular images.

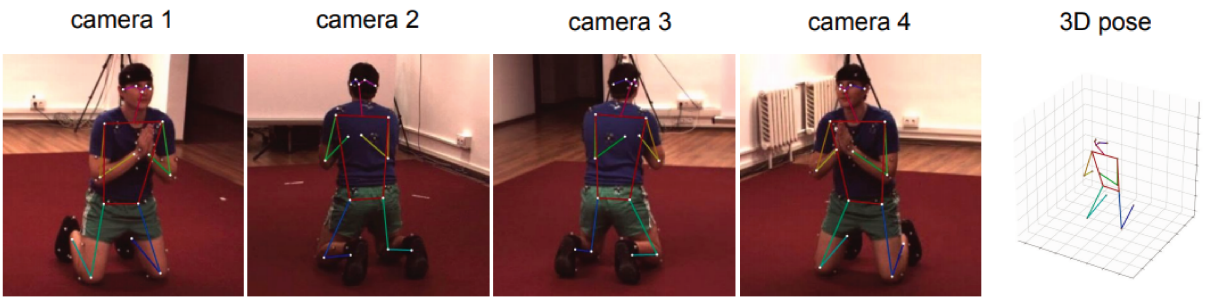


Figure 3.6: 3D pose estimation based on multi-view images.[104]

3.5.2 Infant 3D Pose Estimation

The enormous progress in human 3D pose estimation techniques inspired researchers to use similar approaches in infant 3D pose estimation tasks. Even if the first 3D pose analysis was achieved by reflective markers [54] using infrared cameras, more recent studies focused on fully marker-less methods. Some researchers tried to obtain a 3D keypoint representation using RGB images in addition to depth images obtained from depth cameras. Wu et al. [106] used a Kinect to capture color and depth videos of infant movements. Then, used OpenPose [65] to estimate infants' 2D movements from RGB images. The 3D coordinates of the infants' joints are achieved after combination with corresponding depth images. They introduced a classification method using an evaluation index as a threshold to identify normal and abnormal movements among the video sequences in the MINI-RGBD dataset. They started by calculating a joint motion complexity and a joint motion correlation in addition to small-world networks,

then combined the 3 parameters to end up with an index (real number) able to separate the two populations (Normal vs Abnormal movements). Even if this method used a more interpretable classification method, the validation of the method on a small dataset like MINI-RGBD is uncertain. In addition, as explained before, the later dataset does not give information about the movements represented in the videos (only that it contains infants up to the age of 7 months) so there is a crucial obstacle to performing automatic GMA using these synthetic video sequences. Li et al.[107] used a similar approach with the same pose estimation network (OpenPose) and image acquisition protocol (Depth camera) but with a correction of depth information to solve the problem of matching between RGB and depth images, since depth values captured by depth camera represent the body surface information only [106] and the depth image will map a joint X and Y to the same depth value if one joint is occluded by another. They created an infant body dimension model that consists of the ratios of the half dimensions of the head, neck, shoulders, elbows, wrists, hips, knees, and ankles body parts. These values were added to depth values detected by the camera to correct depth errors. However, a general hassle with methods that use depth cameras is that the use of infrared light-emitting cameras may have a health impact on infants [88], and when dealing with complex infants' poses with a harder view or obstructive gestures, the method may be misidentified [108]. Shivakumar et al. [109] presented a stereoscopic image acquisition protocol for infant pose estimation. They used two cameras, one placed right above the baby and the other positioned on the baby's right side. To obtain depth images, the team used a series of calibration, rectification, and matching operations to get disparity maps and then depth maps. Since the group used only a tracking algorithm, infants were provided with a blue onesie to facilitate the tracking of the torso center using the segmented blue color mask. The optical flow was used for tracking limbs, and a manual selection of regions was needed where the user clicked on a point within the limb region; then, a marker was set using their marker identification method.

Article	Detection method	Infants(Mean age in weeks)	Extracted features	Classification
[47]	Accelerometers	10 (36.3 GA)	166 Kenimatic features	SVM, DT, and
[48]	Accelerometers	10 (36.5 GA)	Kenimatic features	NB, SVM, ar
[49]	Accelerometers	23 (34.42 GA)	32 Kenimatic features	DT
[50]	EMTS	20 (7 CA)	Joints centers and axes	Threshold
[51]	EMTS	67 (13 CA)	Arm movements	Threshold
[52]	Visual markers	22 (4 CA)	8 Kenimatic features	Quadratic disc
[53]	Visual markers	145 (40 CA)	6 Kenimatic features	Threshold
[56]	Frames difference	82 (14 CA)	8 3 Kenimatic features	LR
[57]	Optical flow	82 (14 CA)	Wavelet analysis	SVM
[58]	LDOF	127 (17 CA)	643 Kenimatic features	LR, AdaBoost,
[59]	LDOF	127 (17 CA)	3 Kenimatic features	LR
[61]	CNN + LSTM	-	-	Softmax (7 cl
[62]	Frames difference	21 (-)	4 Kenimatic features	LLGMN
[3]	2D Pose	51 (10 CA)	50 frames poses	SMNN
[85]	2D pose	4 (-)	-	-
[6]	2D pose	31 (40 GA)	FMA, FMS, and CMA	SVM, RF, and

[7]	2D pose	-	38 Kinematic features	NB
[106]	3D pose	12 synthetic	38 Kinematic parameters	Thresho
[8]	3D pose	MINI-RGBD dataset of 12 synthetic infants introduced in GMA c		
[9]	2D pose	SyRIP dataset of synthetic and real images introduced in GMA c		
[107]	3D pose	A 3D infants' pose estimation method introduced in GMA con		
[109]	3D pose	A 3D infants' pose estimation method introduced in GMA con		

Table 3.1: Summary of the reviewed studies on infants' movements detection and c

Abbreviations: AUC, Area Under the Curve. CA, Corrected Age. CMA, Center of Movement Neural Netwrok. CP, Cerebral Palsy. CS, Cramped Synchronized. DBN, Dynamic Bayesian N EMTS Electromagnetic Tracking Systems. FMA, Factor of Movements Area. FMS, Factor Gestational Age. LDA, Linear Discriminant Analysis. LDOF, Large Displacement Optical Flow. L Mixture Network. LR, Logistic Regression. LSTM, Long Short-Term Memory network. NB, Naiv ROC, Receiver Operating Characteristic. SMNN, Shallow Multilayer Neural Network. SVM, S

3.6 Summary and Thesis Objectives

In the previous sections, we reviewed most of the methods used over the last two decades to automatically assess infants' movements (see Table 3.1). All the studies reviewed suggested a process that consists of two main parts. First, they start by detecting the movements of the infants with multiple techniques and extract information about their quality and quantity. Then, they classify these movements based on the GMA or on the infant's outcome in the future. To automatically detect infants' movements, the sensor-based methods are no longer the SOTA methods since they are time-consuming and impractical. Also, traditional computer vision methods have gradually given way to deep learning networks due to the high performance of neural networks in different tasks. Initially, these networks found applications in human pose estimation in a general context and were subsequently fine-tuned for infant pose estimation. However, in order to deliver accurate results, the fine-tuning process needs important amounts of data (infants' images), which remains a crucial obstacle. This challenge is still up to date due to three main reasons. First, it is a very time-consuming task to acquire infants' images in a clinical environment due to their health conditions. In addition, this needs approval and consent from different third parties (health organizations, parents, etc.). Therefore, very few studies were able to enroll a considerable number of infants, as seen in Table 3.1. Second, due to privacy concerns, the existing small datasets collected by different studies cannot be made publicly available, which explains the lack of such kinds of datasets. Third, the few existing publicly available datasets are either synthetic or are collected from the internet, which makes these datasets not representative of preterm infants in a clinical environment. Also, we can notice that recent studies are focusing on 3D analysis more and more. This is a logical progression since the 2D movement analysis does not exploit the overall infants' movement information in space, and information loss can go up to 53% due to dimensionality reduction [110]. In addition, existing infant 2D pose estimation solutions are not very accurate and have not

progressed in tandem with advancements in human pose estimation (HPE) methods. However, existing 3D infant pose estimations are all based on depth cameras, which are infrared light-emitting cameras that may have a health impact on infants [88]

On the medical aspect and movements classification, we have noticed that most of the studies focused on assessing infants' movements at advanced periods of their birth as a target population, mainly assessing the fidgety movements (starting from 9 weeks post-term) and few of them targeted preterm or writhing movements periods which are an earlier sign to identify neurological and developmental disorders. This can be explained by the barriers that can be faced to record the infants' movements before term age because this population is generally still under monitoring in neonatal intensive care units. Therefore, it requires a special video recording setup to enroll these high-risk infants. Moreover, the majority of studies used multiple machine and deep learning methods (Decision Trees, SVM, Logistic regression,..etc.) to classify infants' movements. This kind of classification lacks interpretability and is referred to as "black boxes" since it does not provide a way for clinicians to understand the resulting insights.

Based on all of these observations, this thesis in the AGMA study context addresses the main challenges faced by previous studies as follows: (1) From October 2020 until June 2023, we created an important dataset of preterm infants' videos (282 videos from 129 infants), born before 33 weeks of gestational age (GA) and hospitalized in the Neonatology department of the Centre Hospitalier Universitaire de Saint-Etienne, France. These videos were reviewed and evaluated by an expert group composed of experienced General Movements Trust-certified assessors. (2) Propose the first framework for 3D infant pose estimation from stereoscopic images dedicated to preterm infants hospitalized in the Neonatal Unit. This 3D framework does not use any infrared camera and is totally non-invasive. (3) Provide 2D infant pose estimation models based on SOTA convolutional neural network (CNN) architectures specifically retrained with our dataset images and put at the disposition of the scientific community. (4) Develop a deep-learning-based software to fully automatically select general movements

sequences from long video recordings. (5) Introduce the mean 3D dispersion (M3D) as a quantitative index of preterm movements according to GMA, based on a new normalized 3D representation characterizing the spontaneous activity of infants.

3.7 Conclusion

This chapter reviewed most of the methods used to automatically assess infants' general movements. This review helped to explore the progress made in this field of study and also to identify the actual challenges that need to be addressed. Also, it served as guidance to set the orientations of our research efforts, as described in the summary section.

4

Stereoscopic 3D pose Estimation Framework

Outline

4.1 Introduction

The first step toward automatically assessing the infants' general movements is the automatic detection of these movements. This chapter goes through the details of our approach to automatically estimate the infants' poses in 3D. It starts with introducing the video acquisition protocol used in this study to capture infants' videos and images. These infants' images served as a unique and particular dataset AGMAPose for training and fine-tuning multiple SOTA human pose estimation networks such as HRNet, HigherHRNet, and DarkPose, which are detailed in the Method section. We also explain the experiments performed and the comparison of the results obtained from different architectures, including the ones that are widely used for infant pose estimation nowadays, such as the retrained OpenPose network and FiDIP. Our approach includes two different tasks: 2D pose estimation and 3D pose estimation, so the results and discussion are reported in the same order. This chapter ends with a conclusion which summarizes the significant presented findings.

4.2 Video Acquisition

4.2.1 Video Acquisition Protocol

To automatically assess infants' general movements, a real and important dataset of infants' images is needed. The first step to achieve this goal is to create a suitable video acquisition system that must be safe and non-invasive. In this study, a sophisticated video recording protocol (see Figure 4.1) was put in place in the Neonatology department

of the Centre Hospitalier Universitaire de Saint-Étienne, France. This video recording protocol included a radiant warmer in which infants were put in a supine position wearing a diaper and filmed according to Prechtl's method of GMA [2]. Even if there are no objective studies on the effect of light illuminance on the quality of infants' movements, we have noticed that infants tend to turn their heads away from the light source in case of high illuminance. Therefore, a lux meter was used to control the room light illuminance between 60 and 120 lux to avoid having low-quality images and disturbing the quality of GMs. For video acquisition, A ZED2 (Stereolabs, San Francisco) stereoscopic camera was used with frames of 1280x720 resolution at 30 Frame Per Second (FPS), and a computer to run the ZED2 camera Application Programming Interface (API).

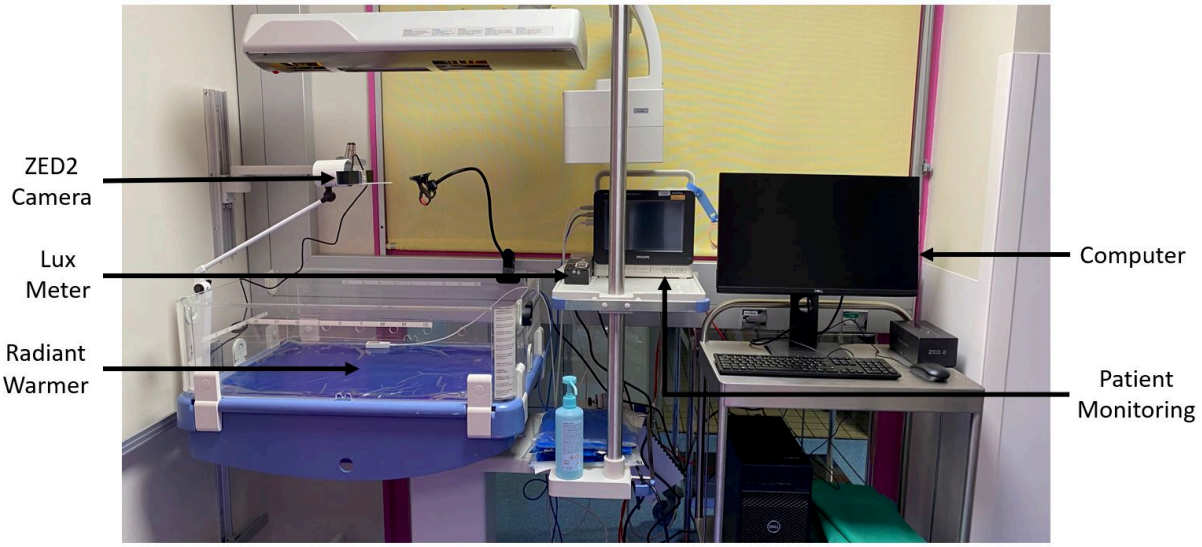


Figure 4.1: Video acquisition protocol.

4.2.2 ZED2 Camera

The ZED 2 stereo camera is a 3D camera that can be used for a variety of applications, including object tracking, 3D mapping, and 3D scene reconstruction. It has 2 high-resolution cameras (1280x720 pixels) spaced at 12cm distance, making it suitable for depth sensing, creating 3D models, and 3D scene reconstruction. The camera API offers access to the ZED2 camera's features. This API allows control of the camera

and captures images and depth data. For depth sensing, the API uses a stereoscopic method to fuse images from the two cameras to create a 3D depth map. The choice of this camera as a video recording tool was based on two considerations. First, the camera offers the ability to capture both 2D and 3D information. This will allow a complete visualization and analysis of infants' movements in space. Second, and unlike existing 3D cameras (Microsoft Kinect,...) that emit infrared light to calculate depth information, the ZED2 camera does not emit any infrared light that may have a health impact on infants [88] thanks to its totally no-invasive stereoscopic system.

4.2.3 Stereo Vision Systems

The human binocular vision can have a three-dimensional perception of the environment through the eyes due to the difference observed in the images from the left and right eyes. These images are not the same, with a slight difference in the position of the objects due to the separation between the eyes; the brain can determine the distance that the objects are. This principle is nowadays used in computer stereo vision to recreate a 3D scene representation based on the two images from different viewing points. First, the two images must be rectified. The goal of this stereo rectification is to reproject image planes onto a common plane parallel to the line between camera centers. Then, a depth map can be obtained from the disparity between the pixels in both views using triangulation. The ZED2 camera has two monocular cameras with a known focal length f and a distance between the two projection planes called baseline $b = 12cm$.

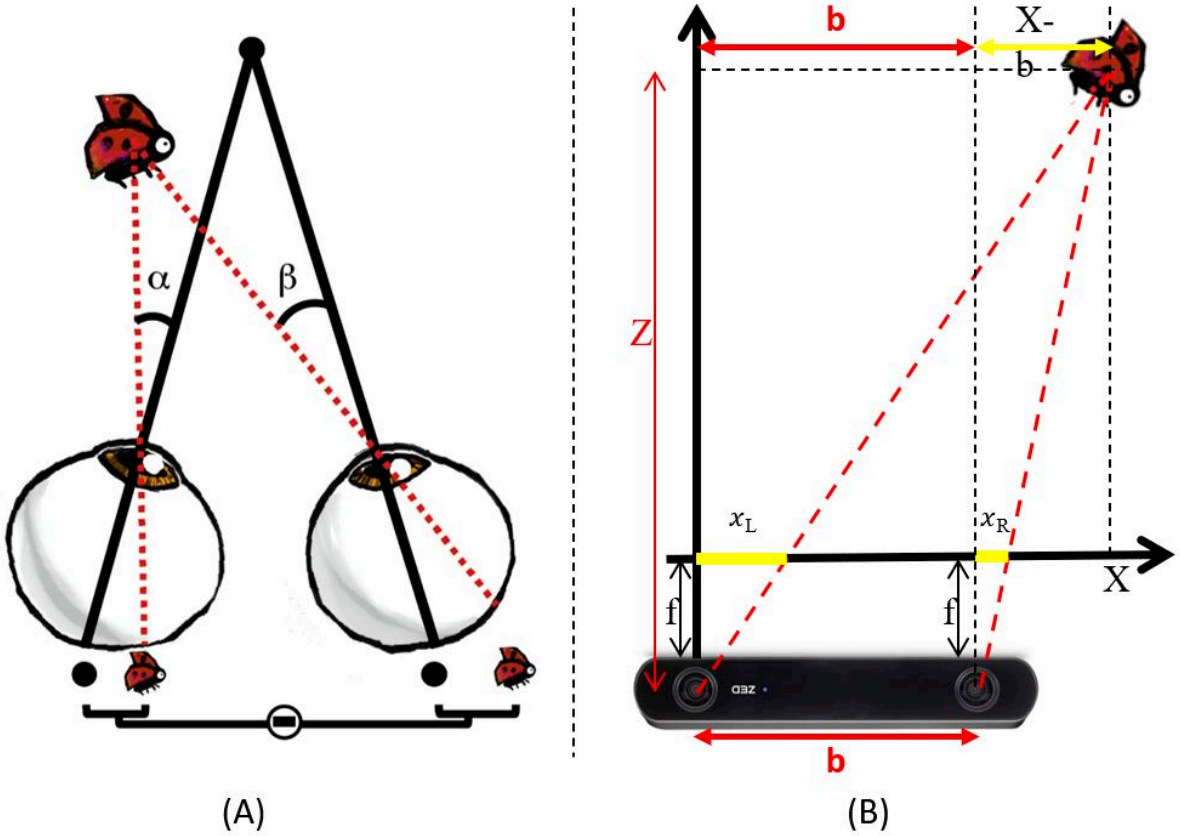


Figure 4.2: Human and ZED2 camera vision systems. (A) Human binocular vision system [111], (B) ZED2 stereo vision system.

Following Figure 4.2(B), any object with real-world coordinates (X, Y, Z) with respect to the left camera will be captured with a slight difference in pixel coordinates on the X-axis between the left view x_L and the right view x_R, this difference is called disparity D with $D = x_L - x_R$. So, according to Thales' theorem:

$$\frac{f}{Z} = \frac{x_L}{X} \quad and \quad \frac{f}{Z} = \frac{x_R}{X - b}$$

therefore:

$$X = \frac{Z x_L}{f} \quad and \quad \frac{f}{Z} = \frac{f x_R}{Z x_L - bf}$$

Hence:

$$Z = \frac{Z x_L - bf}{x_R}$$

So:

$$Z = \frac{bf}{x_L - x_R} = \frac{bf}{D} \tag{4.1}$$

The pinhole camera model describes the mathematical relationship between the real-world coordinates of a point $P(X, Y, Z)$ and its projection onto the image plane $p(x, y)$ as :

$$\begin{cases} x = f_x \times (\frac{X}{Z}) + c_x \\ \\ y = f_y \times (\frac{Y}{Z}) + c_y \end{cases} \tag{4.2}$$

where f_x, f_y, c_x, and c_y are the ZED2 camera intrinsic parameters which are known. Since the depth value Z can be calculated using triangulation, the X and Y values can be calculated as follows:

$$X = [(x - c_x) \times Z] / f_x \quad and \quad Y = [(y - c_y) \times Z] / f_y \tag{4.3}$$

4.3 Dataset

4.3.1 AGMAPose Dataset

The AGMAPose dataset was built by recording real images of 53 premature infants enrolled in the AGMA study (see Appendix A.2), born from October 2020 to June 2021. All infants born before 33 weeks of gestational age (GA) and hospitalized in the Neonatology department of the Centre Hospitalier Universitaire de Saint-Étienne, France, were included with a written parental consent. Exclusion criteria were the ongoing presence of ventilatory support, contraindication of a radiant heat warmer, and absence of written parental consent. The features of the included children are summarized in Table 4.1. Each included infant was recorded from one to three times, with a period ≥ 7 days between two consecutive recordings. The AGMAPose dataset contains 44,250 stereoscopic image pairs (88,500 images in total) from a selection of videos from 53 infants acquired between 32 and 41 weeks of GA (see Figure 4.3). 175

short videos of 5 seconds (30×5 frames) duration were manually selected from long videos ($\geq$ 1-hour) at the moments where infants move the most in order to get different poses. In addition, 10 one-minute videos were automatically selected from other long videos using the frame differencing method (described in section 5.2.1).

Table 4.1: AGMAPose dataset population features. SD: standard deviation.

Male sex, n (%)	31 (58%)
Birth Gestational age, weeks, mean (SD)	30 (0.3)
Birth weight, g, mean (SD)	1357 (40)
Birth weight z-score, mean (SD)	-0.23 (1.02)
Birth length, cm, mean (SD)	38.6 (3.9)
Birth length z-score, mean (SD)	-0.48 (1.01)
Head circumference, cm, mean (SD)	27.7 (2.4)
Head circumference z-score, mean (SD)	0.01 (1.07)

Figure 4.3: A snippet of AGMAPose dataset images collected in a clinical environment.

Ethics

The AGMA study (IDRCB 2020-A03335-34) was approved by the Comité de Protection des Personnes – Sud-Est II Ethical Committee in February 2021. Written parental consent was obtained from each participant. The study was conducted in accordance with international ethical standards and the Declaration of Helsinki.

4.3.2 Dataset Labeling

In order to train or fine-tune any deep learning model, a considerable amount of training data is needed. The human pose estimation networks particularly are supervised learning tasks. So, to train a model to automatically estimate the location of joints in a person's body image, it requires an important amount of images in addition to the ground truth detections. This data (images + ground truth detections) will allow the network to adjust its weights and learn how to precisely estimate the body joints on new images. Hence, AGMAPose dataset images were manually annotated and reviewed to serve as training data. The image labeling process was done according to the COCO keypoint detection task, using 17 points forming the infant skeleton with a bounding box around it (see Figure 4.4). These annotations were performed manually by a group of annotators working for a company that is a leader in data labeling for machine learning and AI applications (Isahit). They were given stereo images containing the two rectified left and right views, and the task consisted of successively pointing each skeleton joint on both views for each image. The consistency of the labeling across time was subsequently checked for each image to guarantee an optimal accuracy of annotations.

4.3.3 Dataset Comparison

The first motivation to create the AGMAPose dataset was the absence of a publicly available and important dataset of infants with real images filmed in a clinical environment. In order to retrain the Openpose network on infant images, Chambers et al. [7]

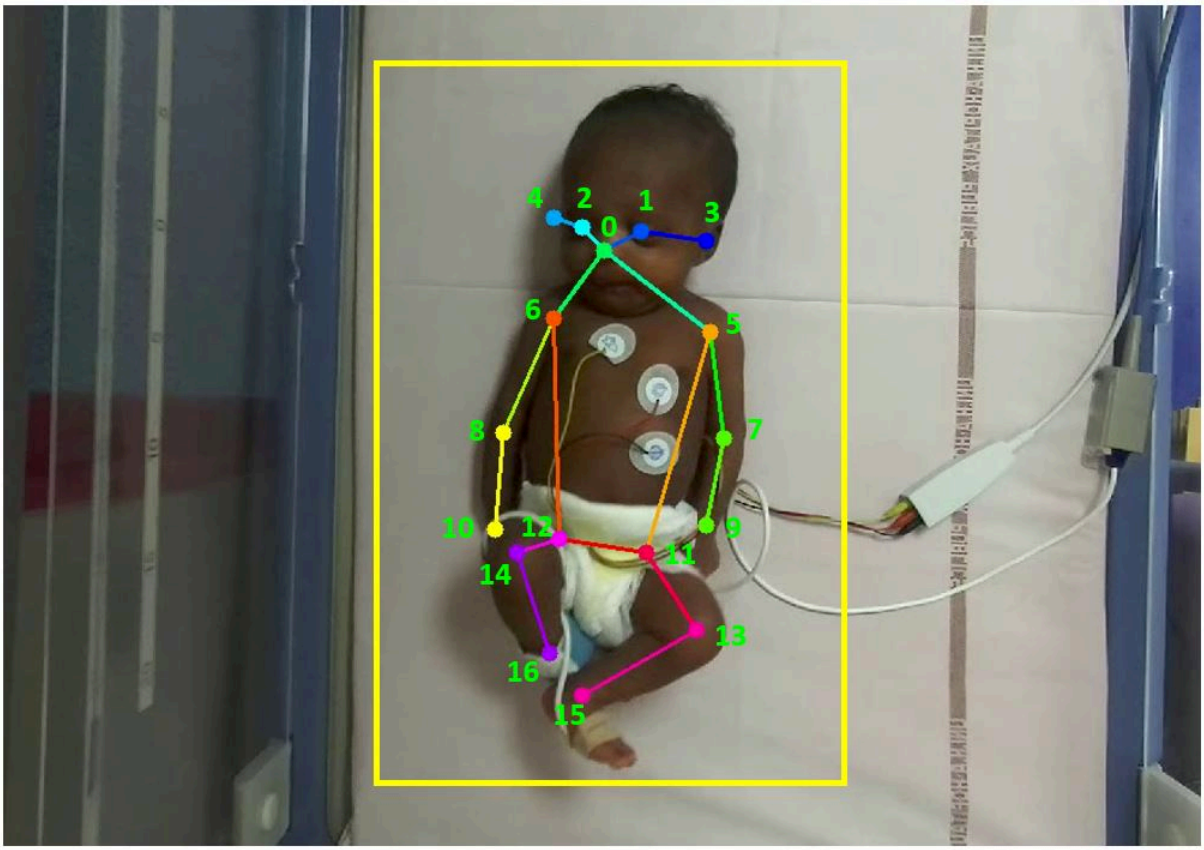


Figure 4.4: Image annotation process: 17 keypoints forming infant skeleton with a bounding box.

collected a dataset mainly from YouTube, which contains images of infants in the wild that are different than images of the targeted population (preterms). The infants were in different positions with too many occlusions. Therefore, when they used the Vatic tool to label the body keypoints for training, it had very poor accuracy, and some joints were not labeled at all as can be seen in their publicly available dataset. The widely used dataset is the MINI-RGBD for Moving INfants In RGB-D [8], which is 100% synthetic. Starting from only 12 RGBD videos of moving infants with 1000 frames each, and for privacy concerns, the team created synthetic 2D infants' bodies. After that, they used a textural mapping to a 3D model called SMIL in order to get 12 movement sequences. These sequences presented very simplistic movements, which makes it poor in terms of the variation of poses. Also, the synthetic nature of these images affected the pose estimation models trained on MINI-RGBD and became unable to generalize well on

Table 4.2: Comparison of AGMA dataset with other publicly available datasets

Dataset	N of images	Type	Source	Context	Age	Condition
MINI-RGBD	12000	Synthetic	SMIL[8]	Children's hospital	0-6 months	Constraint conditions
SyRIP	1700	Synthetic & real	Youtube & Google	Extra hospital	Not available	In the wild
AGMA	88500	Real	Clinical	Neonatology department	32–41 weeks of GA	GMA Protocol

Abbreviations: GA, gestational age; GMA, general movement assessment; N, number. MINI-RGBD [8] stands for Moving INfants In RGB-D dataset and SyRIP [9] for Synthetic and Real Infant Pose dataset.

real-world infants' images [9]. Based on these observations, Huang et al.[9] created the SyRIP dataset, which is a mixture of synthetic and real infant images. Similarly to what Chambers et al.[7] did, they collected 700 2D pictures from YouTube and Google. Then, they used a textural mapping to SMIL 3D model in addition to changing backgrounds, lighting, and camera position to get another 1000 totally synthetic images. Besides being small as a dataset (1700 images), and partially synthetic, the dataset images do not represent the context and poses encountered in the GMA assessment method (see Figure 4.5). Unlike these datasets, the AGMAPose dataset contains images that are 100% real, which is important for retraining deep learning models. In addition, they were captured in a controlled protocol following Prechtl's method and within a clinical setting. Table 4.2 summarizes the characteristics of the three datasets.

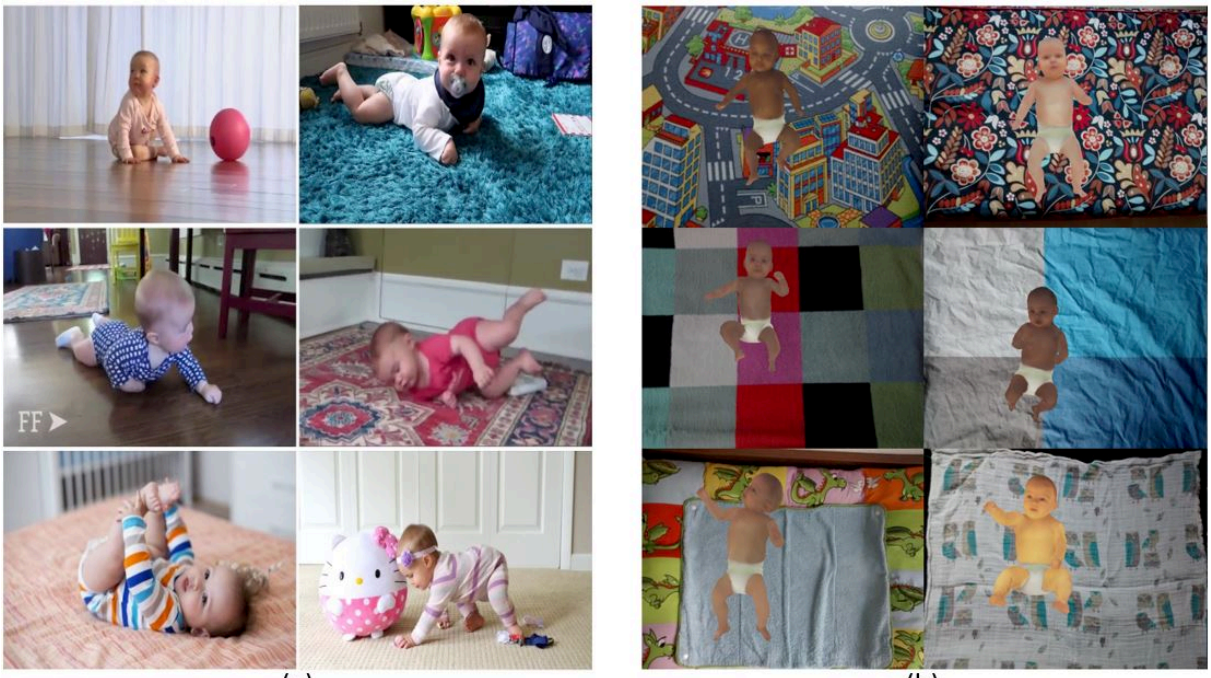


Figure 4.5: SyRIP and MINI-RGBD datasets images. (a): A snippet of SyRIP real images, (b): A snippet of MINI-RGBD images.

4.4 Method

4.4.1 Stereoscopic Pose Estimation Framework

The 3D pose estimation framework presented in this study is based on stereoscopic imaging (see Figure 4.6). As described in section 4.2, the stereoscopic camera captures the same scene from two different views. Therefore, for each rectified view, we estimate the 2D infant poses using a deep-learning model. After getting the pixel coordinates (x, y) of the 17 body joints, we use equations 4.1 and 4.3 to find their real-world coordinates (X, Y, Z).

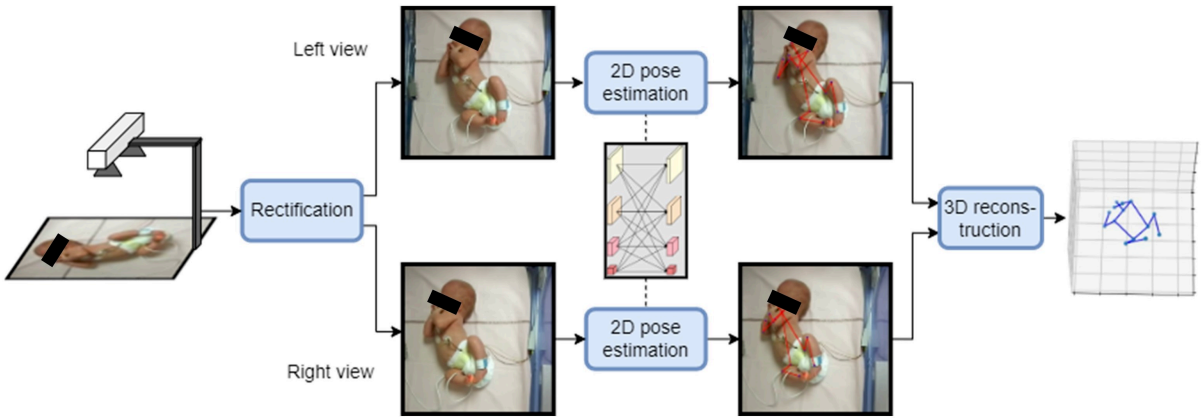


Figure 4.6: 3D infant pose estimation framework based on stereoscopic imaging, 2D pose estimation and 3D reconstruction using triangulation.

Unlike depth maps of RGB-D cameras, which represent the depth information of the visible body surface only, this stereoscopic framework is able to estimate the depth of keypoints even in the case of occlusion. The 2D pose estimation methods have shown strong robustness to occlusions, which makes them beneficial for this task. Taking advantage of this, and using triangulation, the framework can reduce significantly the 3D pose estimation error. Moreover, regarding the infants' health, this framework does not use any infrared light, which can have an impact on infants' eyes, as it is the case with RGB-D cameras.

4.4.2 2D Pose Estimation

The 2D infant pose estimation task consists of automatically estimating the location of infants' body joints (e.g., elbow, wrist, etc.) without the use of any markers. As reviewed in section 3.4, human pose estimation networks are progressively achieving better results on the SOTA datasets such as COCO [63] or MPII [64]. However, these models cannot be directly used for infant pose estimation tasks and can be inaccurate since they were trained on data containing images of adults only who are structurally and anatomically different compared to infants. On the other hand, instead of training these models on infants' images from scratch, a better approach is to fine-tune the models pre-trained on adults' images. This is a transfer learning technique used in deep/machine learning domains in which knowledge learned from a task is re-used in order to boost performance on a related task [112]. In this study, three different deep neural architectures were studied and evaluated. They were originally created for adult human pose estimation based on the High-Resolution Network HRNet [81]: HRNet itself, HigherHRNet, and DarkPose.

HRNet

HRNet was introduced by Ke Sun et al. [81] for different tasks, including human pose estimation, segmentation, and object detection. It became a state-of-the-art top-down pose estimation network due to the strategy of using parallel networks of different resolutions instead of traditional in-series high-to-low networks (Figure 4.7). The network calculates the high-resolution sub-network in parallel with lower-resolution sub-networks. Then, the sub-networks are fused through the fuse layers such that each of the different resolution representations receives information from other parallel representations over and over. Maintaining high-resolution representations throughout the entire network makes the architecture a more robust backbone for computer vision problems [113] and more suitable for infant pose estimation. For this architecture, Ke Sun et al. [81] created two different networks, a small one called HRNet32 (28.5 Million parameters)

and a big one called HRNet48 (63.6 Million parameters). Each network has two different versions with an input image size of 256×192 and 384×288. They were initialized by the weights of the models pre-trained on the ImageNet dataset and then trained on the COCO train2017 dataset, including 57K images and 150K person instances.

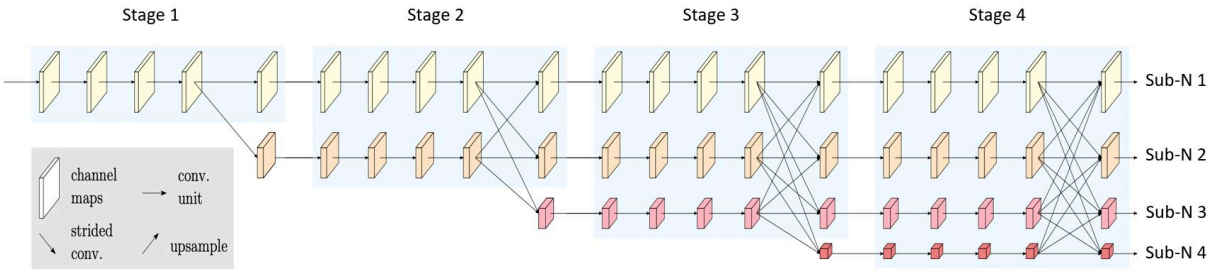


Figure 4.7: HRNet network architecture with 4 stages and 4 subnetworks. Sub-N: Subnetwork. [114]

The HRNet network starts by putting the images (256x192 or 384x288) into a stem, which consists of two stride-2 3×3 convolutions decreasing the resolution to $\frac{1}{4}$ of the initial image resolution (64x48 or 96x72). The output of this stem is used as input to Stage 1 in the main body shown in Figure4.7. The main body starts from a high-resolution convolution stream as the first stage, then gradually adds lower-resolution streams one by one, forming new stages and subnetworks. Each subnetwork repeatedly receives information from other parallel subnetworks through the fuse layers, either by strided convolutions from higher-resolution to lower-resolution subnetworks or by nearest neighbor up-sampling from lower to higher-resolution subnetworks (see Figure 4.8). This method allows lower-resolution branches to capture contextual information and higher-resolution branches to preserve spatial information.

In the end, instead of using the direct keypoint regression where the output is the (x, y) coordinates for each keypoint, HRNet uses a heatmap representation of size 64x48 or 96x72 as output, denoting the likelihood of the joints to be detected in a given section of the input image (see Figure 4.9). This method was first introduced in 2014 by Tompson et al. [115] and rapidly became the most commonly used coordinate representation. It will be discussed more deeply in the next sections.

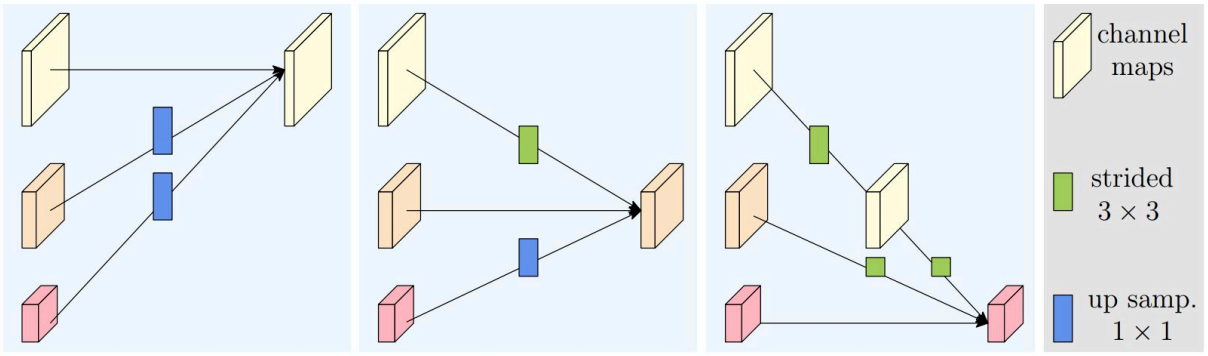


Figure 4.8: Information exchange units between high, medium, and low resolutions subnetworks [114].

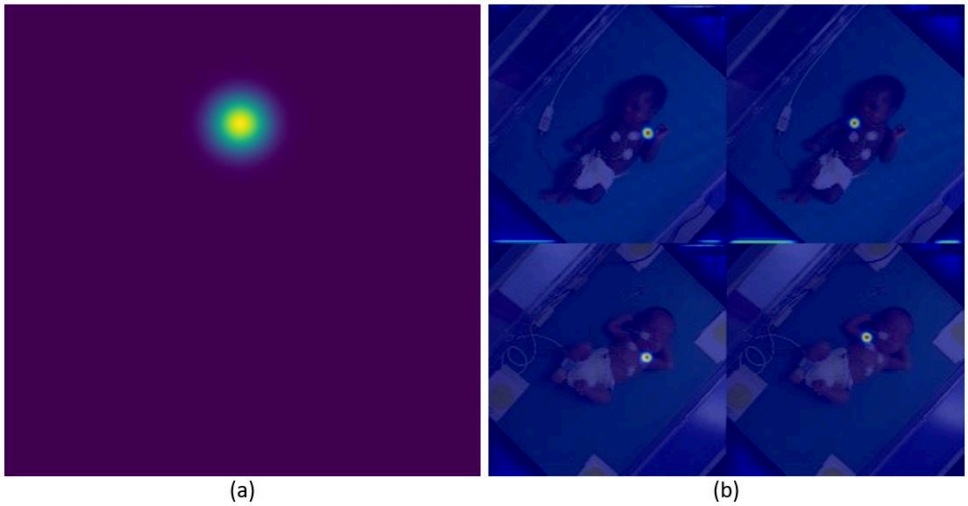


Figure 4.9: Heatmaps representations. (a) Heatmap representing the likelihood of a joint detected in image. (b) Keypoints heatmaps of infants' right and left shoulders.

HigherHRNet

This network [75] uses HRNet as a backbone after the accurate results HRNet had in HPE tasks. The aim of this network was to address the problem of scale variation in bottom-up pose estimation by outputting multi-resolution heatmaps and using the high-resolution representation backbone of HRNet. As shown in Figure 4.10, The HigherHRNet architecture is similar to HRNet but with an additional block to the end for outputting higher-resolution heatmaps. Unlike traditional bottom-up methods that predict heatmaps at $\frac{1}{4}$ the resolution of the input image, and based on the importance of heatmap resolution for an accurate keypoints prediction, HigherHRNet uses an

additional deconvolution module to generate heatmaps 2 times larger in resolution than the input feature maps ($\frac{1}{2}$ the resolution of the input image). This is achieved by a 4×4 deconvolution followed by BatchNorm and ReLU, in addition to 4 residual blocks after deconvolution to refine the upsampled feature maps. This feature pyramid with two resolutions is then used in two ways. In the training stage, HigherHRNet uses multi-resolution supervision, where not only one ground truth heatmap is used but two with different resolutions ($\frac{1}{4}$ and $\frac{1}{2}$). They are created by using a Gaussian kernel (see equation 4.4) with a standard deviation of 2, centered at the ground truth keypoints location. In the inference stage, the network uses bilinear interpolation to upsample all the predicted heatmaps with different resolutions to the resolution of the input image and average the heatmaps from all scales for final prediction.

$$\text{2D Gaussian Kernel:} \quad G(x, y, u, v) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-u)^2+(y-v)^2}{2\sigma^2}} \tag{4.4}$$

where σ is the standard deviation and (u, v) are the keypoint 2D coordinates.

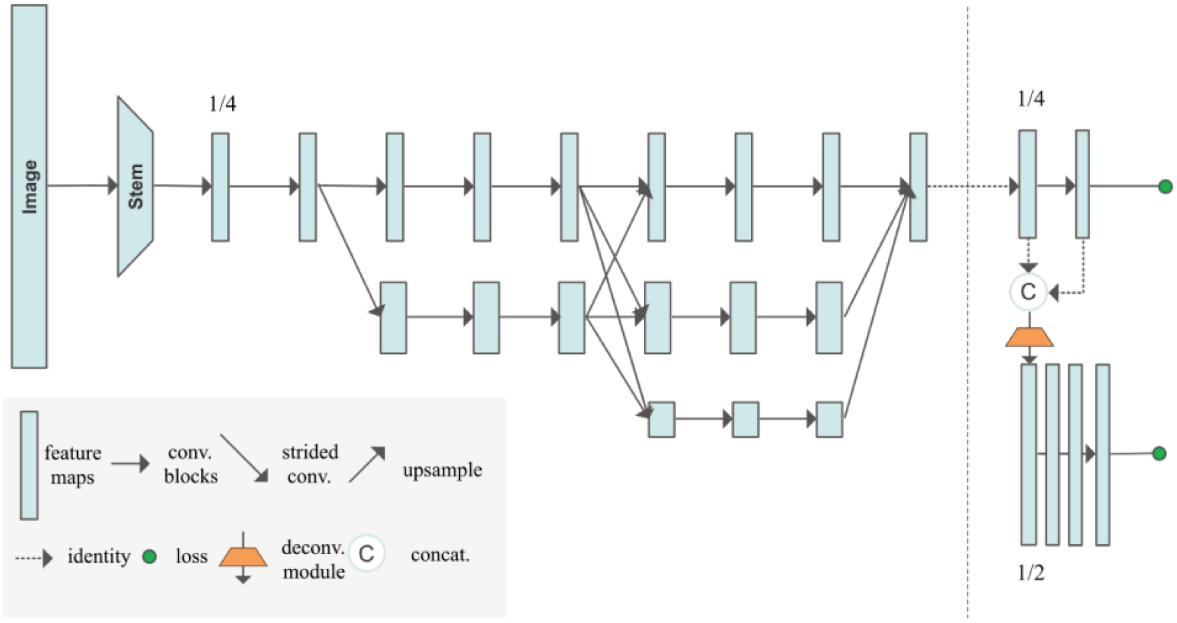


Figure 4.10: HigherHRNet network architecture [75].

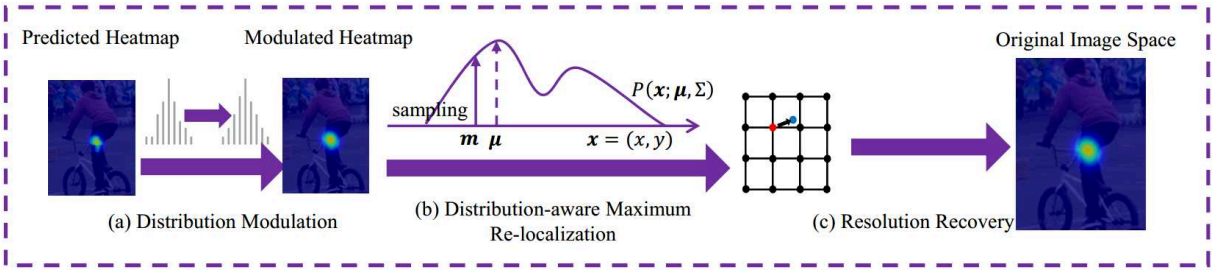


Figure 4.11: Distribution-Aware coordinate Representation of Keypoint (DARK) method [83].

DarkPose

The added value in this architecture is the fact that it can be used as a plugin that improves the performance of SOTA Human pose estimation models. Previously, research studies focused on improving network architectures for more effectively regressing the heatmap supervision, including HRNet and HigherHRNet. However, the founders of DarkPose aimed to investigate the issues of heatmap representation on human pose estimation. To train a model, the labeled ground-truth coordinates of a joint are encoded into a heatmap as a supervised learning target, and when testing, the predicted heatmap output of this model is decoded into the coordinates in the original image coordinate space. In this process, Zhang et al.[83] noticed that the heatmaps predicted by HPE models do not have a good Gaussian structure compared to the training heatmap data but usually present multiple peaks around the maximum activation, which can affect the model performance. Therefore, they suggested starting with a heatmap modulation, where a Gaussian kernel is used to smoothen the predicted heatmap by simple convolution (see equation 4.5).

$$h^{'} = K \circledast h \tag{4.5}$$

where h is the original heatmap, $h^{'}$ the modulated heatmap, K the gaussian kernel, and $\circledast$ convolution operator.

After that, instead of using the standard coordinate decoding method as in equation 4.6, where the coordinate prediction is based on the coordinates of the maximal and

second maximal activations in the predicted heatmap, they proceed with a maximum re-localization with a Taylor expansion to find the maximal activation of the modulated heatmap as in equation 4.7. In the end, since heatmaps do not have the same spatial size as the original image, they upsample the heatmaps to the original image resolution by the inverse of the ratio $\frac{1}{\lambda} \in \mathbb{R}^+$ used in downsampling the input image.

$$p = m + 0.25\frac{s-m}{\|s-m\|_2} \tag{4.6}$$

where m and s are the coordinates of the maximal and second maximal predicted heatmap activations.

$$\mu = m - (D''(m))^{-1}D'(m) \tag{4.7}$$

where D'' and D' are the derivatives of the modulated heatmap, and m is the coordinates of the maximal activation of the predicted heatmap.

4.4.3 Experiments

The goals of this section can be described as follows. First, to compare the different selected deep learning architectures and their retrained versions on our dataset. Second, to test existing 2D pose estimation methods on our data collected in a clinical environment, and compare them with the networks we retrained. Third, quantify the precision and accuracy of the keypoint detection in the perspective of clinical use.

The OpenPose network, which was originally trained on adult human pose estimation, is retrained by Chambers et al.[7] on an infants' dataset of 9039 images collected mainly from YouTube. This retrained network is widely used to automatically assess infants' movements. So, we compare and test this model without further training on our images and check whether it can generalize well on data from a clinical environment only. The DarkPose + FiDIP network [9] was a suggested solution to solve the lack of clinical data. Based on a hybrid dataset of synthetic and real images collected from the internet

(SyRip), a domain classifier was trained to promote a feature extractor to retain the ability to extract keypoints' information but also to ignore the differences between the real and the synthetic input images. This network is also tested and compared without extra training as an existing solution in addition to OpenPose retrained version.

The experiments in [81] have shown that HRNet networks with higher input resolution (384×288) improve the AP compared to networks with lower input resolution (256×192). For that reason, two HRNet pre-trained networks (W32 & W48) are chosen with an input resolution of 384×288, in addition to two HigherHRNet pre-trained networks (W32 & W48) with an input resolution of 640×640, and finally, two DarkPose on top of HRNet pre-trained networks (W32 & W48) with input resolution of 384×288. These networks were tested with and without training on our infants' images clinical dataset. For training, we used only 15 epochs since we were only adjusting the network weights. An Adam optimizer was used, and the learning rate was set to 10^{-4} throughout the whole process. We have used data augmentation methods that can help to increase the amount of the training data considerably. This includes random image rotations between $-45°$ and $45°$ and random image re-scaling (between 0.9 and 1.1). However, unlike Sun et al., we did not use half-body data augmentation since the goal is to get complete infants' poses.

Evaluation metrics

Even if our dataset ground truth annotations were performed according to the COCO keypoint detection task [63], its respective evaluation metric (AP, AP50, AP75) was not used since COCO AP is calculated using the Object Keypoint Similarity (OKS) metric. As shown in equation 4.8, OKS uses per-keypoint constants that are calculated only from images in the COCO validation set [116], which cannot be representative of another kind of dataset.

$$OKS = \frac{\sum_i \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right)\delta(\upsilon_i > 0)}{\sum_i \delta(\upsilon_i > 0)} \tag{4.8}$$

where d_i is the Euclidean distance between the detected keypoint and the corresponding ground truth, v_i is the visibility flag of the ground truth, s is the object scale, and k is a per-keypoint constant that controls falloff.

The Percentage of Correct Key-points (Percentage of Correct Key-points (PCK)) is more suitable for infants' 2D pose estimation evaluation. It defines a correctly detected joint if the distance between the predicted joint location and the ground truth is smaller than a certain threshold defined by a fraction of torso size. For this work, the torso height (the distance between the right hip and right shoulder) was used instead of torso diagonal size (the distance between the right hip and left shoulder) since torso diagonal size can vary depending on the infant's pose as can be seen in Figure 4.13. The notations PCK@0.2, PCK@0.1, and PCK@0.05 were used to refer to the fractions of torso height used as thresholds 0.2, 0.1, and 0.05 respectively (4.10).

$$PCK@X = \frac{\sum_{i=1}^{N} \delta(d_i^2 \leq X \times T)}{N} \tag{4.9}$$

where d_i is the Euclidean distance between the detected keypoint and the corresponding ground truth, X is the threshold fraction, T is the torso height and N the total number of keypoints.

For 3D pose evaluation, 3DPCK was used. It defines the keypoint as correct if it falls within a given distance from the ground truth: 1cm, 2.5 cm, 5cm, and 10cm. These thresholds can show how accurate the keypoints estimations are at different levels. For clinical usage and considering that the average infant height is 50 centimeters, a 2.5 cm error will represent 5% of the body size which is acceptable. But errors that exceed this value (5cm, 10cm) are considered as too important. On the other hand, errors that do not exceed 1cm can be tolerated for this use case.

$$3DPCK@X = \frac{\sum_{i=1}^{N} \delta(d_i \leq X)}{N} \tag{4.10}$$

where d_i is the distance between the detected keypoint and the corresponding ground truth in centimeters, X is the threshold distance in centimeters, and N the total number of keypoints.

4.5 Results and Discussion

4.5.1 2D Infant Pose Estimation Results

Figure 4.12 shows the results obtained after training and testing the different architectures. The error is calculated in pixels as the Euclidean distance between the ground truth and predicted keypoints and averaged for each image (36k images with 880x720 pixels resolution). The impact of retraining the networks on our infants' dataset is obvious, as it led to a noticeable decrease in the mean error across all architectures, as well as a reduction in the standard deviation. Specifically, for the HigherHRNet32 and HigherHRNet48 networks, the mean error diminished from 30.7 to 3.7 pixels and from 28.5 to 5.3 pixels respectively.

The two networks OpenPose and FiDIP, have been evaluated without any additional training. With a confidence score threshold of 0.1, it was observed that OpenPose exhibited the lowest mean error of 6 pixels (see Fig. 4.12). This result is remarkably low; however, it should be noted that the mean error calculation exclusively considered successfully predicted joints across the entire dataset and the corresponding mean detection rate for this network was approximately 5 keypoints per image, in contrast to the other networks that achieved a detection rate of 17 keypoints per image (100% of joints detected). This discrepancy becomes evident when assessing the network's performance using PCK evaluation at various thresholds, as illustrated in Table 4.3. In fact, OpenPose performed poorly in comparison to networks that were not trained on any infants' images.

As referred in Table 4.3, the FiDIP network exhibited a comparable PCK when compared to the non-retrained DarkPose48 network (95.35% versus 95.32% for PCK@0.2).

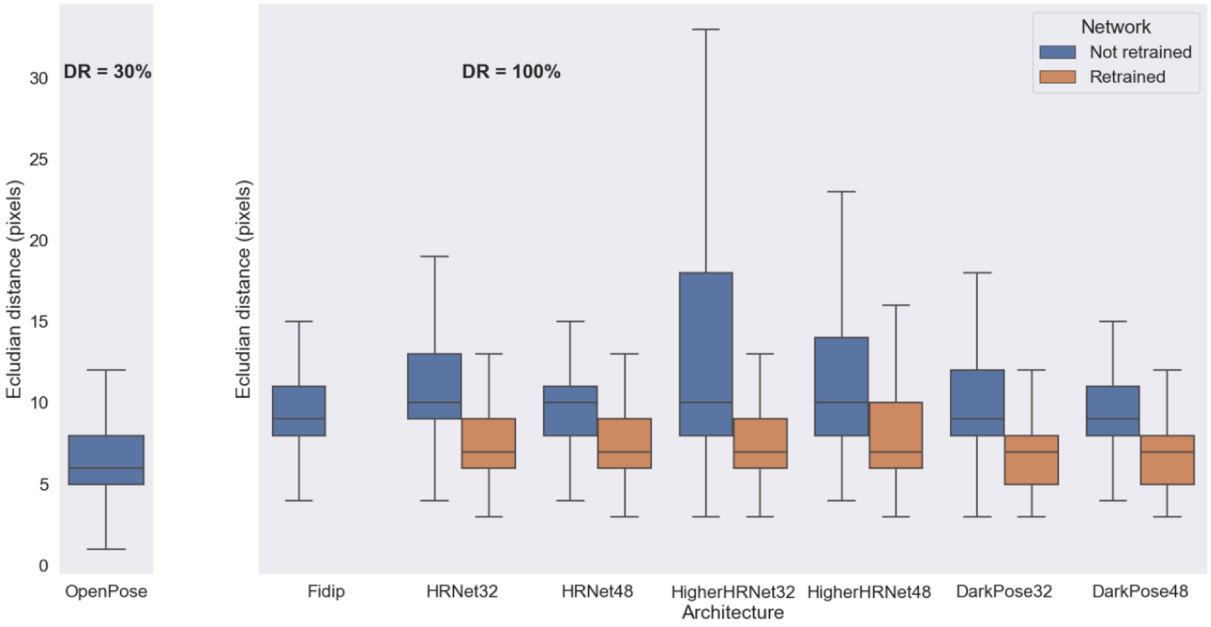


Figure 4.12: 2D pose estimation error for different networks in pixels. The boxplot shows the median error, and the whiskers mark 1.5 times the interquartile range. DR: detection rate.

Table 4.3: 2D pose estimation results comparison of original *versus* retrained networks on our test dataset.

	Original			Retrained		
Network	PCK@0.05	PCK@0.1	PCK@0.2	PCK@0.05	PCK@0.1	PCK@0.2
Openpose	50.08	60.31	64.07	-	-	-
FiDIP	55.93	84.29	**95.35**	-	-	-
HRNet32	52.31	79.95	92.41	66.35	90.92	97.98
HRNet48	55.86	83.49	95.01	65.84	90.82	98.25
Higher32	53.80	75.58	85.24	65.90	89.47	97.44
Higher48	54.89	77.07	87.78	65.16	88.27	96.41
DarkPose32	54.44	82.09	93.31	**68.07**	**91.37**	**98.30**
DarkPose48	**56.58**	**84.39**	95.32	66.93	90.97	98.28

Despite being retrained using the SyRIP dataset, which comprises 700 real and 1000 synthetic images, the FiDIP network did not demonstrate any advantage over DarkPose48. This lack of improvement can be attributed to the characteristics of the images within this dataset (refer to Fig 4.5 and Table 4.2), which were sourced from the internet and contain infants' poses that do not adequately represent the typical video acquisition protocol for assessing infants. Consequently, retraining on synthetic images does not

inherently enhance the efficiency of networks when processing real images captured in a clinical setting. This underscores the need for a real and significant dataset in this particular research domain, which cannot be overlooked or substituted.

The HigherHRNet networks demonstrated a satisfactory level of performance in terms of PCK but not as good as HRNet networks. This disparity can be attributed to the fact that the scale of our subjects within the image is already adequate. Consequently, upsampling the predicted heatmaps and averaging them across all scales may potentially impact the accuracy of the predictions. This observation aligns with the findings of Cheng et al. [75], who concluded that HigherHRNet performs exceptionally well when dealing with small scales. Furthermore, it is worth noting that the HigherHRNet32 network exhibited a notable standard deviation of error. Upon closer examination, it was identified that in certain test images, the network exhibited confusion between keypoints located on the left and right sides. However, this issue was successfully addressed through network retraining, as it was no longer observed thereafter.

Both architectures of DarkPose (W32 & W48) achieved the best PCK (91.37% & 90.97% at 0.1 threshold and 98.30% & 98.28% at 0.2 threshold respectively). This proves that instead of using the standard coordinate decoding method, the heatmap decoding process described in 4.4.2 can improve infants' pose accuracy, which is in accordance with the results in [83]. Another observation is that after training the W32 and W48 versions, they have approximately the same performance, and this is because our subject's bounding boxes are large enough, so a lighter version of these networks can be used to reduce time complexity and resources.

4.5.2 3D Infant Pose Estimation Results

The 3D pose estimation results were in accordance with the results obtained in 2D analysis (see Table 4.4). Openpose had the least 3DPCK at all thresholds with the same detection rate of 30%, and FiDIP had comparable results with not retrained DarkPose48 (93.88% versus 93.86% at 3DPCK@5cm) despite being retrained using the SyRIP

dataset, which validates that training on artificial images alone does not automatically improve the effectiveness of networks.

All the retrained networks gained 5% minimum of 3DPCK@2.5cm, and their mean error was reduced. The retrained version of DarkPose32 achieved the best 3DPCK at all thresholds, with a minimum mean error of 1.72 cm, which is a very promising result regarding the complex poses in our testing dataset. The results for DarkPose48 network after training were very similar to the 32 version with only a 3mm difference in mean error.

Moreover, when evaluating the results of this latter network for each group of joints separately, the mean error can be even lower, particularly for keypoints that are easy to define, such as the nose for which the retrained DarkPose32 network achieved 99.74% in 3DPCK@2.5cm and a mean error of 0.94 cm (see table 4.5). The same table shows how higher the error is for keypoints that are difficult to localize precisely, such as the hips, for which the mean error is more than 2cm. We can also notice that better 2D pose detections lead to better 3D pose estimations, as is the case for DarkPose Networks, and the opposite is valid also since HigherHRNet had a considerable standard deviation and mean error.

Figure 4.13 shows the results of some networks on different complex infants' poses. It can be observed that all three networks before training do not provide accurate 3D pose estimations, which become better after training even when joints are half or completely occluded, as is the case for the second image. A standard depth camera will not be able to estimate the depth of these joints since it represents visible body surface depth information only, which shows the benefits and advantages of using our framework to analyze infants' movements.

In summary, the main conclusions that can be drawn from these results are: (1) it is particularly important to use a dedicated annotated dataset to train pose detection models for infants in a real medical environment; (2) compared to existing retrained models Openpose and FiDIP, models built on the HRNet network provide better precision, particularly DarkPose32 which is the best performing one in our case; (3) the latter

Table 4.4: 3D pose estimation results comparison of the non-retrained networks on our test dataset.

Original	Openpose	FiDIP	HRNet32	HRNet48	Higher32	Higher48	DarkPose32	DarkPose48
3DPCK@1cm	17.46	32.71	26.46	28.72	31.87	32.63	31.81	**32.97**
3DPCK@2.5cm	40.75	76.26	67.47	72.31	65.94	68.06	73.64	**76.46**
3DPCK@5cm	53.20	**93.88**	88.63	92.56	81.03	84.08	91.20	93.86
3DPCK@10cm	57.12	**98.38**	95.80	98.00	87.94	90.12	96.96	98.20
Mean error (cm)	2.44	**2.31**	3.77	2.62	8.96	8.52	2.97	2.37
STD (cm)	**5.60**	8.06	24.39	6.91	70.62	72.18	16.81	8.81
Retrained	-	-	HRNet32*	HRNet48*	Higher32*	Higher48*	DarkPose32*	DarkPose48*
3DPCK@1cm	-	-	35.43	35.33	39.33	37.56	**40.49**	40.24
3DPCK@2.5cm	-	-	79.47	79.99	80.48	78.35	**83.11**	82.92
3DPCK@5cm	-	-	96.66	96.74	95.75	93.93	**97.29**	97.26
3DPCK@10cm	-	-	99.33	99.43	98.69	97.79	**99.45**	**99.45**
Mean error (cm)	-	-	1.91	1.90	1.98	2.36	**1.72**	1.73
STD (cm)	-	-	2.50	4.48	7.34	10.39	**2.31**	2.34

* Retrained on AGMAPose dataset.

Table 4.5: Per joints 3D pose estimation results of the retrained DarkPose32 network.

	Nose	Shoulders	Elbows	Wrists	Hips	Knees	Ankles
3DPCK@1cm	65.38	27.57	42.19	48.70	15.12	44.34	35.36
3DPCK@2.5cm	95.41	81.61	86.78	87.28	66.93	88.68	82.99
3DPCK@5cm	99.74	98.25	98.55	98.07	95.08	98.76	98.00
3DPCK@10cm	99.99	99.98	99.88	99.58	99.94	99.93	99.57
Mean error (cm)	0.94	1.72	1.45	1.42	2.24	1.39	1.71
STD (cm)	0.96	1.11	1.14	1.59	1.42	1.07	3.73

network provides a 3D localization error of 1.7cm with a 3DPCK at 2.5cm of 83% and a 3DPCK at 5cm of 97%.

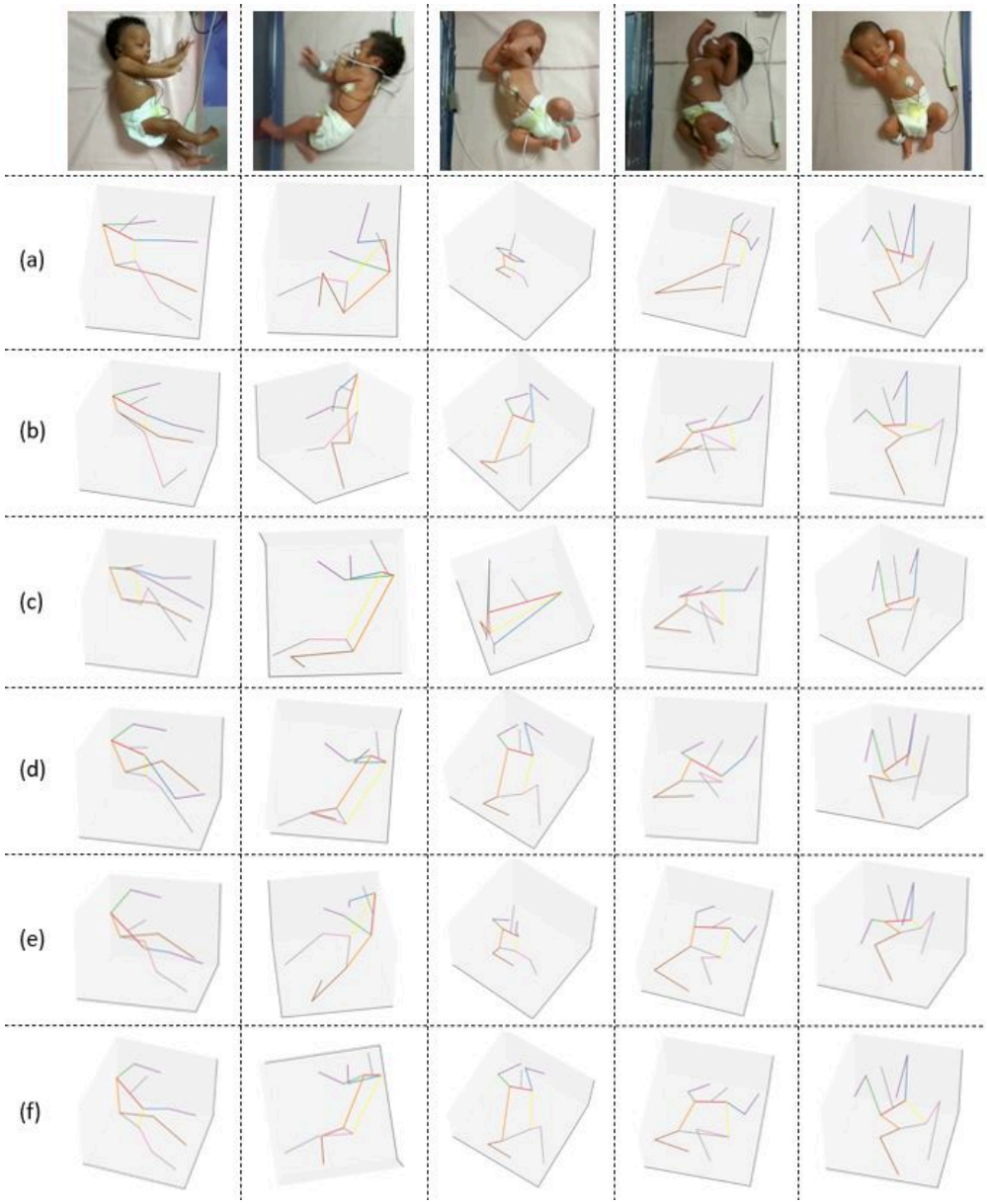


Figure 4.13: 3D pose estimation results of different networks. (a) HRNet32, (b) retrained HRNet32, (c) HigherHRNet32, (d) retrained HigherHRNet32, (e) DarkPose32, (f) retrained DarkPose32.

4.6 Conclusion

This chapter has introduced the stereoscopic 3D infants' pose estimation framework used in this study. Compared to other existing automatic infant movement analysis tools, the AGMA stereoscopic framework has successfully shown the possibility of estimating accurate 3D infants' poses without the use of any markers or infrared cameras. Three state-of-the-art 2D human pose estimation networks (HRNet, HigherHRNet, and DarkPose) were retrained on a dataset of 88,500 preterm infants' images collected in a real medical environment and manually annotated. The networks were tested with 18000 stereoscopic images and compared to the latest works on 2D infant pose estimation retrained on real or synthetic images. The minimum mean 3D error on joint position achieved was 1.7 cm with a 3DPCK at 5 cm of 97%. The study demonstrated that retraining on synthetic images does not consequently make networks efficient on real images captured in a clinical environment and that an adapted real dataset was more advantageous. In addition, it showed that an appropriate heatmap decoding process could improve infants' pose accuracy instead of using the standard coordinate decoding method. The advantage of the presented framework is that any other advanced deep neural network can be used in the future for 2D pose estimation and then 3D pose reconstruction. Beyond GMA, AGMA 3D framework could also pave the way for the development of new tools based on the analysis of preterm infant movements, such as the automation of clinical seizure detection [117] and sleep quantification [118]. However, the next step after having successfully shown the accuracy and the reliability of our framework, is to proceed to the automatic qualitative and quantitative GMs analysis.

5

Automatic General Movements Assessment

Outline

5.1 Introduction

In order to make the GMA easier in clinical practice, this chapter introduces new methods to automatically select GM sequences from long videos, based on frame differencing and pose estimation network retrained previously. The latter was used to create a software to automatically select GM sequences with a user-friendly graphical user interface. Also, this software was used to create the AGMA dataset that contains 282 one-minute videos of GMs from 129 infants born <33 weeks GA. The videos were classified by an expert group based on the GMA method [2]. This large and labeled dataset was used in the automatic assessment of general movements. The method section presents a new classification parameter called Mean 3D Dispersion used to discriminate normal and abnormal GMA based on their complexity more specifically. This classification parameter is based on the stereoscopic 3D framework presented previously, in addition to a new spherical representation and normalization method. It was used to classify infants' movements with the expert group observations serving as ground truths, and the classification performance was measured with different methods as shown in the experiments and results sections.

5.2 AGMA Automatic Sequence Selection

The general movement assessment method relies on recording videos of infants' movements for a duration ranging from 30 to 60 minutes [2]. Then, clinicians must manually extract about three GMs sequences for an accurate assessment. This task includes

visualizing all the videos and trying to manually select 1 to 3-minute video sequences where the infants are moving the most. This time-consuming process becomes an obstacle for clinicians to practice the necessary assessments. Moreover, in order to examine the consistency or inconsistency of normal or abnormal assessment findings, this process needs to be repeated multiple times to get a developmental trajectory for each infant [2]. Therefore, automatic methods for selecting GMs sequences from long videos are needed. In this study, we have used two different automatic methods to select GMs sequences in a specific chronological order. The first one is a frame differencing method used before achieving the results described in section 4.5.1, and the second one is a deep learning-based method derived from those results.

5.2.1 Frame Differencing Method

The frame differencing method is a technique used in computer vision and image processing to detect changes or motion within a sequence of images or video frames. It works by comparing consecutive frames and identifying areas where pixel values have changed significantly. This method is widely employed in various applications, including surveillance, motion detection, and object tracking. The method used in this study can be described as follows: First, every two consecutive RGB video frames (F_i, F_{i+1}) are converted to grey-scale to increase computational speed, then subtracted pixel by pixel. The idea is to subtract the pixel values in one frame from the corresponding pixel values in the next frame. This subtraction results in a new image known as the difference frame. After that, we use a threshold $0 \leq T \leq 255$ to detect the pixel changes between the two frames. If the pixel value is smaller than the threshold (no pixel change means no motion), it is set to 0; otherwise, it is set to 1. This way, we get binary motion images (see Figure 5.1), and once the area of the withe pixels (pixels with value 1) is calculated in each binary image, it results in a time series M_i of values representing the quantity of motion along the video. Therefore, the task of finding the starting frame x of the most important one-minute sequence of GMs from a video of 30

FPS can be described mathematically as finding the starting index of the most important sub-series in M_i, which can be accomplished by a summing window function f of size 60 seconds (1800 frames) as in equation 5.1.

$$x = \underset{j}{\operatorname{argmax}}\, f(j) = \underset{j}{\operatorname{argmax}}(\sum_{i=j}^{j+1800} M_i) \tag{5.1}$$

Even if this method is simple to implement and relatively efficient for real-time applications, it remains very sensitive to noise and changes in lighting. Also, the choice of an appropriate threshold value $0 \leq T \leq 255$ is important because a threshold that is too low may lead to false positives, while a threshold that is too high may result in missed detections.

Figure 5.1: Motion image extraction. (a) RGB color frame. (b) Greyscale color frame. (c) Binary motion image.

5.2.2 Deep Learning-Based Method

This method was created as a solution to the limitations faced with the frame differencing method. It is based on a different approach where infants' motion in videos is not detected using the whole frame pixels as data, but only the infants' body keypoints. Inspired

by the results reported on the infant 2D pose estimation task, this method uses a convolutional neural network for infants' 2D pose estimation. It can be described in two steps. First, using the DarkPose32 network that we have fine-tuned (see section 4.5.1) on AGMA.1 dataset, we estimate the infant pose (17 keypoints) $P = [p_1, p2, ..., p17]$ through all the video frames. Next, for every two consecutive poses P_i and P_{i+1}, we calculate the sum of the Euclidean distances between respective keypoints as in equation 5.2.

$$S_i = \sum_{k=1}^{17} \|P_{i+1,k} - P_{i,k}\| \tag{5.2}$$

S_i is a time series that represents the quantity of motion along the video (see Figure 5.2) in the same manner as M_i in the frame differencing method but with more representative and accurate values. Therefore, the task of finding the starting frame x of the most important one-minute sequence of GMs from a video of 30 FPS can be described mathematically as finding the most important sub-series in S_i, which can be accomplished by a summing window function g of size 60 seconds (1800 frames) as in equation 5.3. It can be noticed that differently than the frame differencing method, this method does not depend on a manually chosen threshold to detect infants' movements, which makes it more robust and accurate.

$$x = \underset{j}{\operatorname{argmax}}\, g(j) = \underset{j}{\operatorname{argmax}}(\sum_{i=j}^{j+1800} S_i) \tag{5.3}$$

Deep Learning-Based Sequence Selector Software

The deep learning-based method for automatic sequence selection significantly reduces the time and effort needed by clinicians to assess infants's general movements. Hence, to simplify the use of this method for medical professionals, we created a software that automatically selects the best sequences of general movements of preterm infants. This software can be used in two different languages, English and French (see Figure 5.3. With a simple Graphical User Interface (GUI), The user can upload a video from

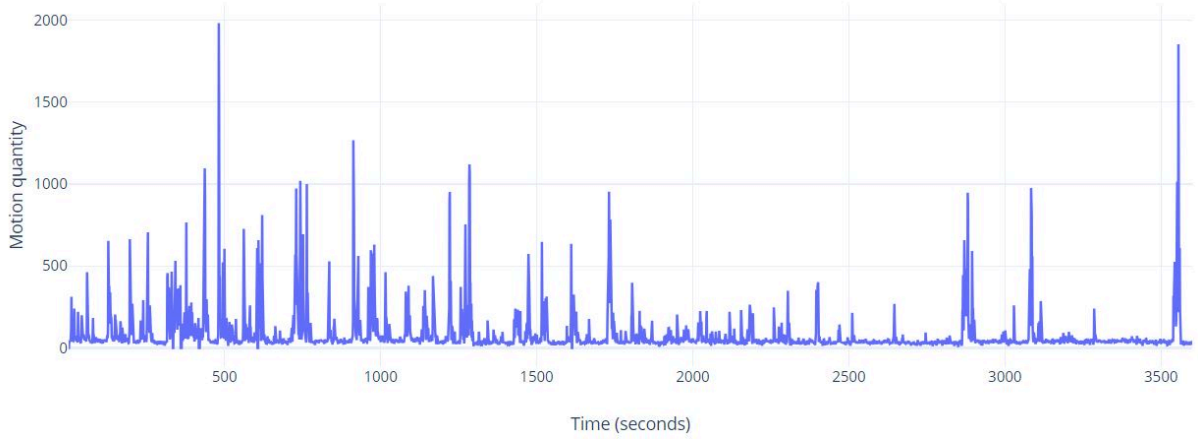


Figure 5.2: Motion quantity values calculated using the deep learning-based method in an infant video of one-hour duration.

his computer and manually define the number of small video sequences wanted, in addition to their duration in minutes. When the start button is clicked, a neural network estimates the infant pose for one image per second instead of all video images (30 FPS) as described in section 5.2.2. This optimization reduces the processing time and the resources needed for the video selection; therefore, it would be easy to use it on computers that are not equipped with efficient GPUs (Graphical Processing Units). After estimating the infant pose, the software selects the video sequences where movements are present the most using equations 5.2 and 5.3. Finally, these sequences are cut and saved in descending order and according to the user preferences. During the whole process, a real-time graph shows the change in motion, and the user can stop the video processing at any moment with the stop button (see Figure 5.4). This GUI was developed using Python3.7 programming language and based on Pyqt5. PyQt5 is a group of Python bindings for the Qt application framework, which is a C++ framework for developing cross-platform applications. So PyQt5 allows access to the functionalities of this framework and to create graphical user interfaces in Python. PyQt5 includes QtDesigner, which is a visual tool to design GUIs by dragging and dropping customizable widgets such as buttons, text fields, labels, windows, etc. Then a Python code was developed to connect user interface events (signals) to Python functions (slots).

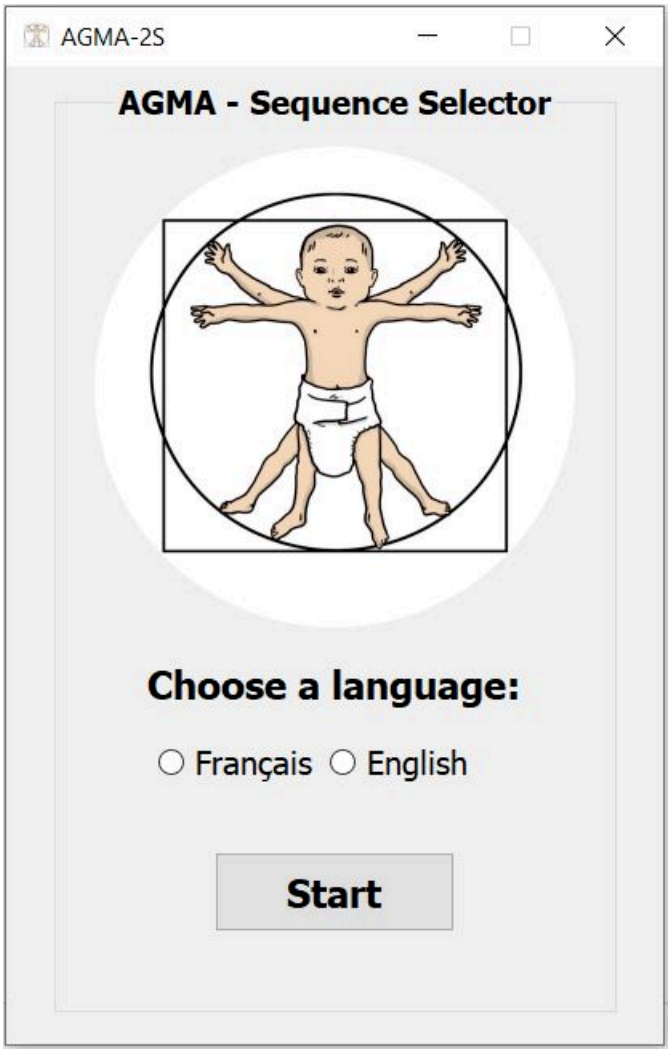


Figure 5.3: AGMA Sequence Selector first window.

5.3 Dataset

5.3.1 AGMA Dataset

Prechtl's general movements assessment method relies on the visual analysis of short videos of infants' spontaneous movements. Therefore, we have created the AGMA dataset that contains 282 videos of one-minute duration from 129 infants (73 male, 56 female). In addition to the 53 infants enrolled for creating the AGMAPose dataset (see section 4.3.1), another 76 infants born from June 2021 to April 2023 before 33 weeks of gestational age (GA) were included. Exclusion criteria were the ongoing

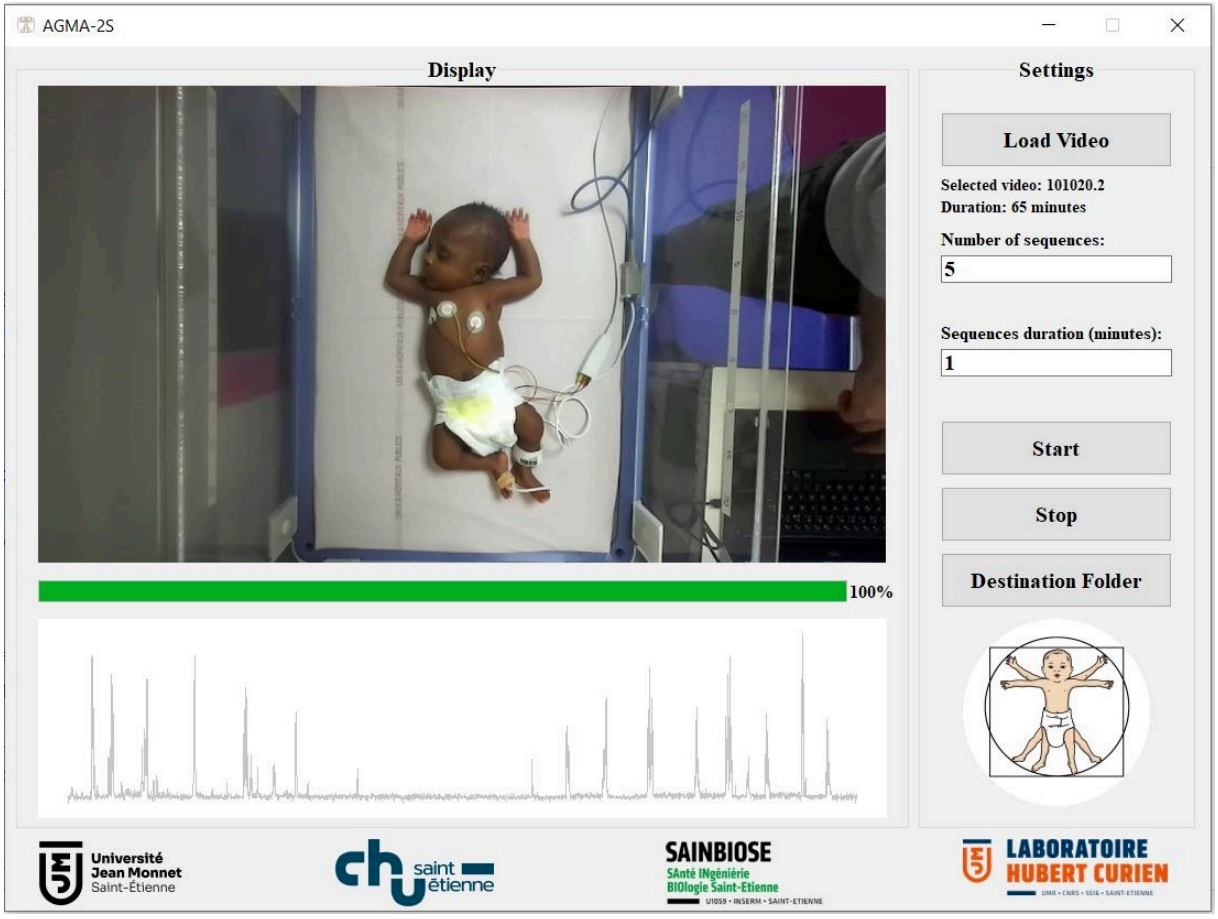


Figure 5.4: AGMA Sequence Selector main window.

presence of ventilatory support, contraindication of a radiant heat warmer, and absence of written parental consent. The features of the whole children enrolled in the dataset are summarized in Table 5.1. Each included infant was recorded from one to three times, with a period ≥ 7 days between two consecutive recordings. The video recordings were performed using the same acquisition protocol described in section 4.2 with an average duration of one hour, depending on the state of the recorded infant. When infants are in a deep sleep, and the spontaneous movements do not appear frequently, we extend the video recording until we observe multiple sequences of GMs. Also, in case of prolonged episodes of fussing, crying, or distraction, the recording is stopped. Since the acquisition camera captures videos from two views, we only select the left-view videos for visualization but keep both views for later processing (explained in the Method

section). The one-hour videos from the left view were processed with the automatic sequence selection methods described in section 5.2 to extract 3 one-minute duration videos at least. These sequences represent the moments where GMs appeared the most and are used for assessing the infants' movements based on GMA method.

Table 5.1: AGMA dataset population features. SD: standard deviation.

Male sex, n (%)	73 (56%)
Birth Gestational age, weeks, mean (SD)	29.8 (2.3)
Birth weight, g, mean (SD)	1324 (403)
Birth weight z-score, mean (SD)	0 (1)
Birth length, cm, mean (SD)	38.3 (5)
Birth length z-score, mean (SD)	0 (0.7)
Head circumference, cm, mean (SD)	27.1 (2.5)
Head circumference z-score, mean (SD)	0 (1)
Age at recording, weeks, mean (SD)	36.7 (2.2)

5.3.2 Dataset labeling

In order to test and validate any supervised classification method, ground truth data annotation is needed. The same applies to the automatic classification of general movements assessment methods. Therefore, all the videos in AGMA dataset were reviewed and evaluated by a GMA-certified group of assessors. From October 2020 to September 2023, monthly group meetings were organized to visualize and assess the video sequences of infants' spontaneous movements recorded during the respective month. These meetings included each time more than 4 certified assessors together, who accomplished the GMA training delivered by the General Movements Trust organization. The meetings started with a calibration session where typical normal preterm and writhing movements were observed in order to train the gestalt visual perception of movements. Then, the recorded videos were projected one by one. At the end of each video, the group discusses about the quality of the spontaneous movements based on their complexity, variability, and fluidity. If one of these three characteristics

is absent, these movements are classified as abnormal. In case of non-agreement between assessors, the videos are played back multiple times until a final agreement. However, for videos where the team could not make a certain judgment mainly for the lack or the total absence of GMs in the video recording, these sequences were classified as non-interpretative. To obtain developmental trajectories and verify the consistency of the assessments' findings for each infant, the GMA Individual Development Trajectory document [2] was used (see Figure 5.6). This document was used to note the assessment classification results as normal, poor repertoire, chaotic, or cramped synchronized. However, through all the 129 infants we have assessed, only normal or poor repertoire movements were observed. The assessment results are as follows: 177 video sequences were evaluated as normal GMs with a mean gestational age of 36 weeks at recording time, 96 video sequences were evaluated as poor GMs (abnormal) with a mean gestational age of 37 weeks at recording time, and 9 non-interpretative video sequences. It should be noted that the same infant can have a different evaluation between two consecutive videos at one-week intervals.

The movements assessment results were either normal or abnormal. However, from the 96 videos evaluated as poor GMs, 50 videos were further classified. The complexity, variability, and fluidity of the movements in these 50 sequences were evaluated, and the decisions were recorded. Therefore, in addition to the main classification classes normal (177 videos) and abnormal (96 videos), the videos can be classified as complex (205) versus not-complex (22), variable (202) and not-variable (25), fluid (200) not-fluid (207) (see Figure 5.5).

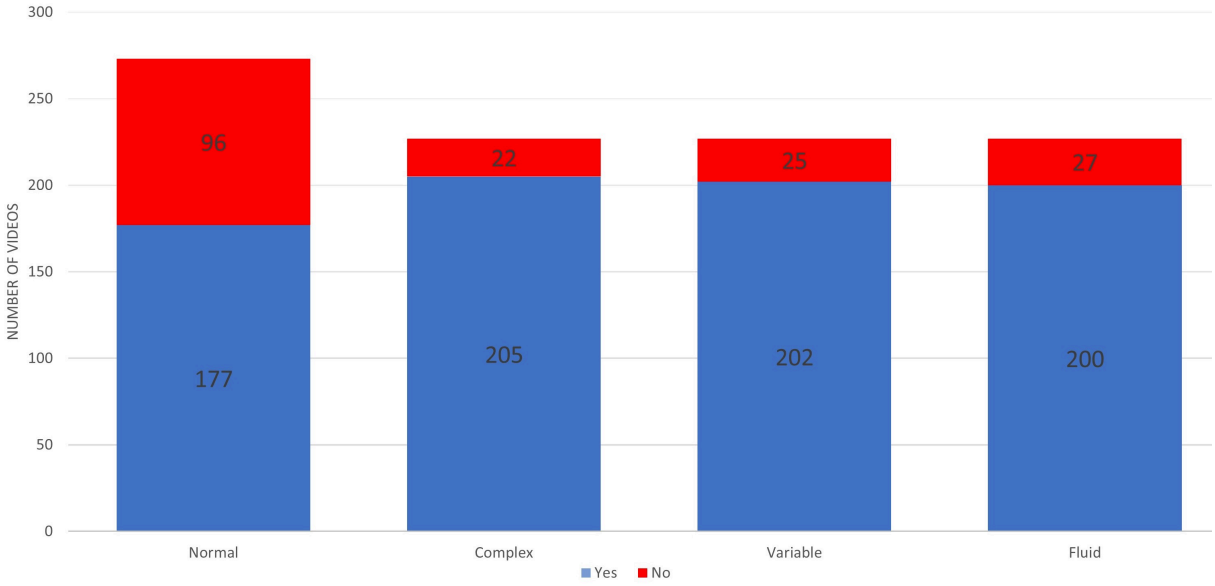


Figure 5.5: Movements classification classes

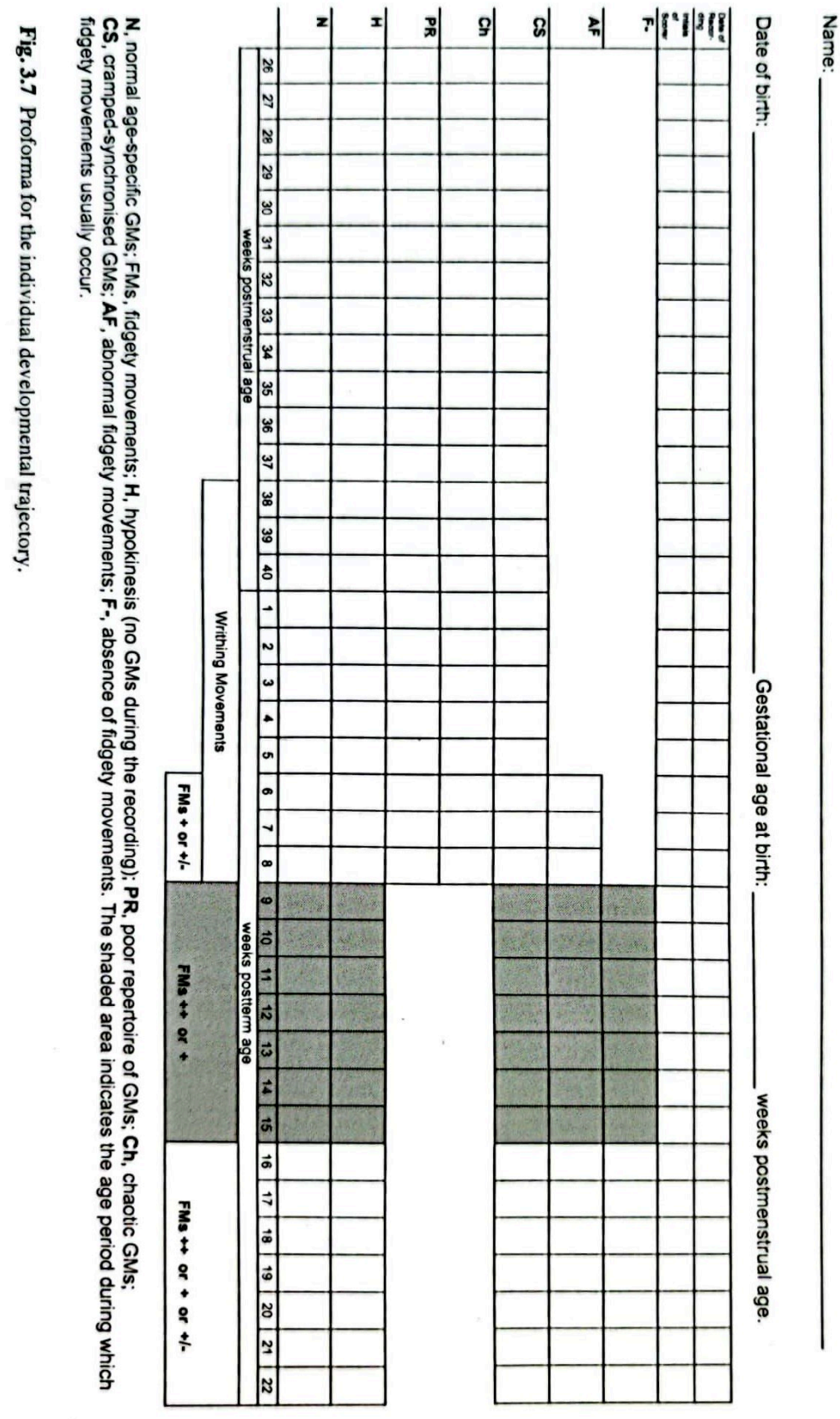


Fig. 3.7 Proforma for the individual developmental trajectory.

Figure 5.6: GMA Individual Development Trajectory [2].

5.4 Method

The automatic GMA method used in this study consists of four main steps. The 3D pose estimation framework presented earlier was used to estimate the infants' 3D pose along the one-minute videos. Then, instead of analyzing the movements with respect to camera origin, new local origins on the infants' bodies were used in order to avoid misrepresentations due to the infants' postures. After that, for an objective comparison of infants' movements with different heights and sizes, we introduce a new spherical representation where each keypoints movement direction is modeled as a point on a unit sphere. This representation allows an objective comparison of the infants' general movements. Finally, and using on the previous representation, the Mean 3D Dispersion parameter is introduced as a classification index of infants' spontaneous movements based on their complexity.

5.4.1 3D Pose estimation and Representation

For all the one-minute duration videos in AGMA dataset, the infants' 3D poses were estimated using the 3D stereoscopic system described in section 4.4.1. Left and right view frames were fed as input to the DarkPose32 network retrained on the AGMAPose dataset to get the 2D pose estimations, then used triangulation to reconstruct the 3D poses in space representing the trajectories of the 17 body keypoints. Analyzing the whole infant's movements in a video means taking into consideration the displacement of the upper and lower body members. Thus, in our study, we specifically analyzed the movement of the elbows, wrists, knees, and ankles. After getting the full 3D trajectories of these body joints, a median filter with a kernel size of 3 was used a long the temporal dimension to get rid of non-accurate points and to correct misidentified joints. Then, in order to analyze these filtered trajectories, we used a method consisting of two steps. The first is changing the relative origin of these 3D trajectories, and the second is a new spherical representation.

Origin Transformation

The 3D estimation process generates real-world coordinates of all keypoints in centimeters with respect to the left camera in the stereo-pair. Consequently, the same infant movements with different baby postures (lying on the back or side) can be considered different since their position relative to the camera is different. Therefore, a new dynamic origin is needed. To address this issue, we used a new origin for each joint (8 in total). These origins are fixed to the infant body in a manner that whatever the position of the infant, similar movements will have similar trajectories and representations.

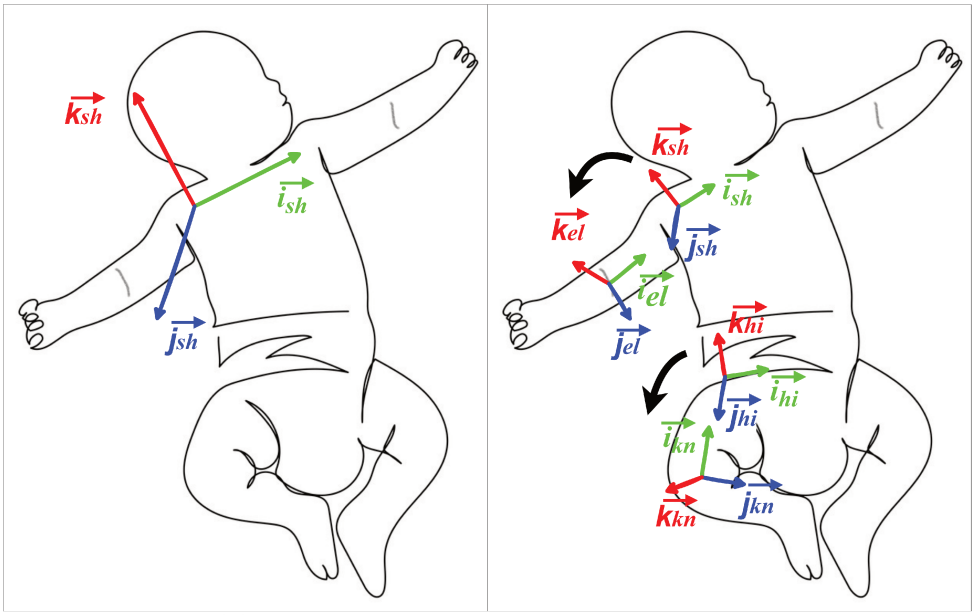


Figure 5.7: Origin Transformation. (a): Local origin for right elbow movement obtained using camera origin rotation and translation. (b): Transformation of shoulder and hip origins to elbow and knee origins.

- For the elbow movements, the shoulder on the respective side of the body was used as an origin. The vector $\vec{i_{sh}}$ connecting the two shoulders was used as the x-axis (abscissa). A cross-product between the two vectors linking the two shoulders with the point in the middle of the hips led to a second vector $\vec{k_{sh}}$ perpendicular to the infant's body plan and representing the z-axis (applicate). Lastly, the y-axis (ordinate) vector $\vec{j_{sh}}$ was obtained as the cross product of $\vec{i_{sh}}$ and vector $\vec{k_{sh}}$ after

normalizing them (see Fig. 5.7). We refer to this new coordinate system by ($\vec{i_{sh}}, \vec{j_{sh}}, \vec{k_{sh}}$) for later usage. Therefore, transforming the real-world coordinates $p_{el} = (x_{el}, y_{el}, z_{el})$ of the elbow points with respect to the camera origin ($\vec{i}, \vec{j}, \vec{k}$) to the new coordinates $p'_{el} = (x'_{el}, y'_{el}, z'_{el})$ with respect to the new coordinate system ($\vec{i_{sh}}, \vec{j_{sh}}, \vec{k_{sh}}$), can be modeled as a rotation of the old origin to the new one, in addition to a translation from the camera origin coordinates $(0, 0, 0)$ to the shoulder coordinates (x_{sh}, y_{sh}, z_{sh}) (see equation 5.4).

$$p'_{el} = R_1.p_{el} + T \tag{5.4}$$

With R_1 the rotation matrix and T the translation vector :

$$R_1 = [\vec{i_{sh}}\ \vec{j_{sh}}\ \vec{k_{sh}}] = \begin{bmatrix} x_{\vec{i_{sh}}} & x_{\vec{j_{sh}}} & x_{\vec{k_{sh}}} \\ y_{\vec{i_{sh}}} & y_{\vec{j_{sh}}} & y_{\vec{k_{sh}}} \\ z_{\vec{i_{sh}}} & z_{\vec{j_{sh}}} & z_{\vec{k_{sh}}} \end{bmatrix} \quad and \quad T = \begin{bmatrix} x_{sh} & y_{sh} & z_{sh} \end{bmatrix}^T$$

so:

$$\begin{bmatrix} x'_{el} \\ y'_{el} \\ z'_{el} \end{bmatrix} = \begin{bmatrix} x_{\vec{i_{sh}}} & x_{\vec{j_{sh}}} & x_{\vec{k_{sh}}} \\ y_{\vec{i_{sh}}} & y_{\vec{j_{sh}}} & y_{\vec{k_{sh}}} \\ z_{\vec{i_{sh}}} & z_{\vec{j_{sh}}} & z_{\vec{k_{sh}}} \end{bmatrix} \begin{bmatrix} x_{el} \\ y_{el} \\ z_{el} \end{bmatrix} + \begin{bmatrix} x_{sh} \\ y_{sh} \\ z_{sh} \end{bmatrix}$$

- For the wrist movements, the elbow on the respective side of the body was used as a new origin ($\vec{i_{el}}, \vec{j_{el}}, \vec{k_{el}}$). The normalized vector $\vec{i_{el}}$ between the elbow and the shoulder on the same side was used as the x-axis. Thus, to get the other two vectors $\vec{j_{el}}$ and $\vec{k_{el}}$, we rotated and translated the previous coordinate system ($\vec{i_{sh}}, \vec{j_{sh}}, \vec{k_{sh}}$) (see Figure 5.7.b). It started by finding the rotation axe $\vec{i_r}(x_r, y_r, z_r)$ as the cross product of $\vec{i_{sh}}$ and $\vec{i_{el}}$. Then the rotation angle $\theta = \arccos\left(\vec{i_{sh}} \cdot \vec{i_{el}}\right)$. Finally, calculated the remaining new origin system vectors as $\vec{j_{el}} = R_2.\vec{j_{sh}}$ and $\vec{k_{el}} = R_2.\vec{k_{sh}}$ where R_2 is the Rodrigues' rotation matrix (see equation 5.5).

$$R_2 = \begin{bmatrix} \cos(\theta) + x_r^2 c' & x_r y_r c' - z_r \sin(\theta) & y_r \sin(\theta) + x_r z_r c' \\ z_r \sin(\theta) + x_r y_r c' & \cos(\theta) + y_r^2 c' & y_r z_r c' - x_r \sin(\theta) \\ x_r z_r c' - y_r \sin(\theta) & x_r \sin(\theta) + y_r z_r c' & \cos(\theta) + z_r^2 c' \end{bmatrix} \quad (5.5)$$

with $c' = 1 - \cos\theta$.

Once the vectors $(\vec{i_{el}}, \vec{j_{el}}, \vec{k_{el}})$ of the new coordinates system are defined, we transform the real-world coordinates $p_{wr} = (x_{wr}, y_{wr}, z_{wr})$ of the wrist points with respect to the camera origin $(\vec{i}, \vec{j}, \vec{k})$ to the new coordinates $p'_{wr} = (x'_{wr}, y'_{wr}, z'_{wr})$ with respect to the new coordinates system $(\vec{i_{el}}, \vec{j_{el}}, \vec{k_{el}})$, by using the same method defined in equation 5.4 as follows:

$$\begin{bmatrix} x'_{wr} \\ y'_{wr} \\ z'_{wr} \end{bmatrix} = \begin{bmatrix} x_{\vec{i_{el}}} & x_{\vec{j_{el}}} & x_{\vec{k_{el}}} \\ y_{\vec{i_{el}}} & y_{\vec{j_{el}}} & y_{\vec{k_{el}}} \\ z_{\vec{i_{el}}} & z_{\vec{j_{el}}} & z_{\vec{k_{el}}} \end{bmatrix} \begin{bmatrix} x_{wr} \\ y_{wr} \\ z_{wr} \end{bmatrix} + \begin{bmatrix} x_{el} \\ y_{el} \\ z_{el} \end{bmatrix}$$

- For the knee movements, the hip on the respective side of the body was used as a new origin $(\vec{i_{hi}}, \vec{j_{hi}}, \vec{k_{hi}})$. The normalized vector $\vec{i_{hi}}$ connecting the two hips was used as the x-axis. A cross product between the two vectors linking the hips by the point in the middle of the shoulders led to a second vector $\vec{k_{hi}}$ perpendicular to the infant's body plan representing the z-axis. Lastly, the y-axis vector $\vec{j_{hi}}$ was obtained as the cross-product of the $\vec{i_{hi}}$ and $\vec{k_{hi}}$ vectors after normalizing them (see Figure 5.7.b). After that, the equation 5.4 was used to transform the real world coordinates $p_{kn} = (x_{kn}, y_{kn}, z_{kn})$ of the knee points with respect to the camera origin $(\vec{i}, \vec{j}, \vec{k})$ to the new coordinates $p'_{kn} = (x'_{kn}, y'_{kn}, z'_{kn})$ with respect to the new coordinates system $(\vec{i_{hi}}, \vec{j_{hi}}, \vec{k_{hi}})$.
- For ankle movements, the knee on the respective side of the body was used as a new origin $(\vec{i_{kn}}, \vec{j_{kn}}, \vec{k_{kn}})$. The normalized vector $\vec{i_{kn}}$ between the knee and the hip on the same side was used as the x-axis. Finally, to get the other two vectors, we rotated and translated the hip origin as we did for the shoulder origin using Rodigues' rotation matrix (equation 5.5) and used equation 5.4 to transform the real world coordinates $p_{an} = (x_{an}, y_{an}, z_{an})$ of the ankle points with respect to the

camera origin $(\vec{i}, \vec{j}, \vec{k})$ to the new coordinates $p'_{an} = (x'_{an}, y'_{an}, z'_{an})$ with respect to the new coordinates system $(\vec{i_{kn}}, \vec{j_{kn}}, \vec{k_{kn}})$ (see Figure 5.7.b).

5.4.2 Spherical Representation

For an objective comparison of the quality of the general movements of multiple infants, we need to eliminate the effect of varying height and body size between infants of different ages and to preserve the angles of the limb's movement. Hence, a new spherical representation was used. For each joint, and after getting the new local coordinates throughout the one-minute videos using the method described in the previous section, we end up with a point cloud representing the movements of the joint with respect to another joint on the same limb as the local origin. Therefore, the joint movement had the shape of a moving point on a sphere with a fixed radius equal to the limb length (see Figure 5.8.a). The advantage of this representation is that it can easily be normalized to a unit sphere, Thus, an objective comparison of the quality of the general movements would be possible. To quantify how complex and multi-directional the infants' movements were in space, we analyze the distribution of joint directions in spherical coordinates by taking the support of this distribution, which means the repetition of the same type of movements by the infant will be represented only once. This was achieved by three steps; the first is converting the points Cartesian coordinates (x, y, z) to spherical coordinates (θ, φ, r) with the radius of unit sphere r = 1 as in equation 5.6.

$$\begin{cases} r = 1 \\ \theta = \arctan\left(\frac{y}{x}\right) if \geq 0; \text{ otherwise } \arctan\left(\frac{y}{x}\right) + 2\pi \\ \varphi = \arccos\left(z\right) \end{cases} \tag{5.6}$$

This conversion will result in a distribution of points with two variables $0 \leq \theta < 2\pi$ and $0 \leq \varphi \leq \pi$. Then, the support of this distribution was projected on an activation map with 128x64 dimensions ranging from $0 \rightarrow 2\pi$ and $0 \rightarrow \pi$ as shown in figure 5.8.c. Therefore, for every point with spherical coordinates (θ, φ), the respective cell (u, v) in

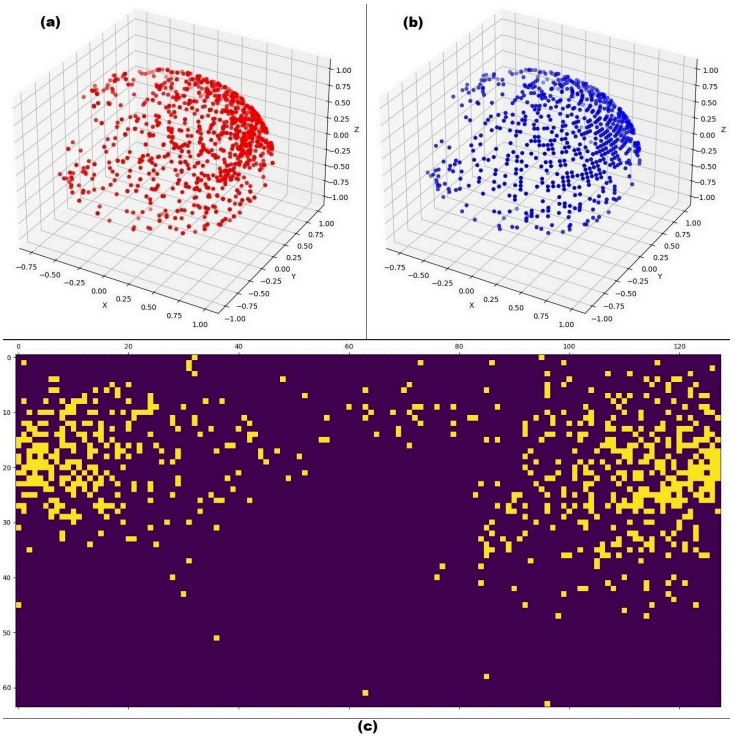


Figure 5.8: Unit sphere representation. (a): The unit sphere describing the movements of the infant's left hand around the elbow before filtering. (b): The unit sphere obtained after filtering. (c): The activation map resulted from the points of the sphere (a).

the activation map was set to 1 such that $\theta_{u-1} \leq \theta < \theta_u$ with $\left(\theta_u = \frac{\pi u}{64}, u = 1, 2, .., 128\right)$, and $\varphi_{v-1} \leq \varphi < \varphi_v$ with $\left(\varphi_v = \frac{\pi v}{64}, v = 1, 2, ..., 64\right)$.

At the end, the activated cells were converted back to Cartesian coordinates, resulting in filtered unit spheres (see Fig. 5.8.b) without point duplications and noisy detections and ready to be used for statistical analysis.

5.4.3 Mean 3D Dispersion parameter (M3D)

To quantify how complex and multi-directional the infants' movements were in the space with a simple and interpretable method, the Mean 3D dispersion parameter was used. For all the 273 videos recorded, and for each joint, the 3D dispersion parameter was calculated using the filtered unit spheres representation described earlier, as follows: given a spherical point distribution with N points, each point has an associated unit direction vector $\vec{v_i}$ (where $i = 1, ..., N$) representing its Cartesian coordinates (x, y, z).

These vectors can reveal important information about the quality of the movement of each joint. If the movements are complex and go in different directions, the direction vectors will be dispersed around the sphere. On the other hand, if the movements are not complex and have a repetitive character, they will be concentrated and polarized in one region of the sphere. Based on this observation, we introduced the 3D polarization parameter that describes the degree of alignment of the direction vectors and calculated as the magnitude of the sum of all vectors $\vec{v_i}$ (where $i = 1, ..., N$) divided by the number of these vectors as in equation 5.7.

$$p = \frac{1}{N} \left| \sum_{i=1}^{N} \vec{v_i} \right| \tag{5.7}$$

The polarization value, as can be deduced from equation 5.7, will always be between 0 and 1. Points that are concentrated in one particular direction, as in Figure 5.9, will have an important polarization since the magnitude of their vector sum will be high, and inversely, the points that are well distributed all around the sphere will have low polarization since the magnitude of their vector sum will be low. Hence, knowing that $0 < p < 1$, the 3D dispersion parameter $\sigma = 1 - p$ describes how well the points are scattered around the origin, denoting that a normal joint movement that is complex and goes in all directions will have a significant 3D dispersion compared to a movement that is repetitive and poor. The advantage of this classification parameter σ is its simplicity and interpretability compared to other classification methods, which is a key parameter in GMA.

5.4.4 Experiments

The two populations (normal and abnormal GMA) videos in AGMA dataset were analyzed and evaluated by calculating the M3D parameter of each joint. Then, for every infant, these parameters were averaged to get one general value of M3D that characterizes the quality of the infants' movements. Based on the M3D values calculated,

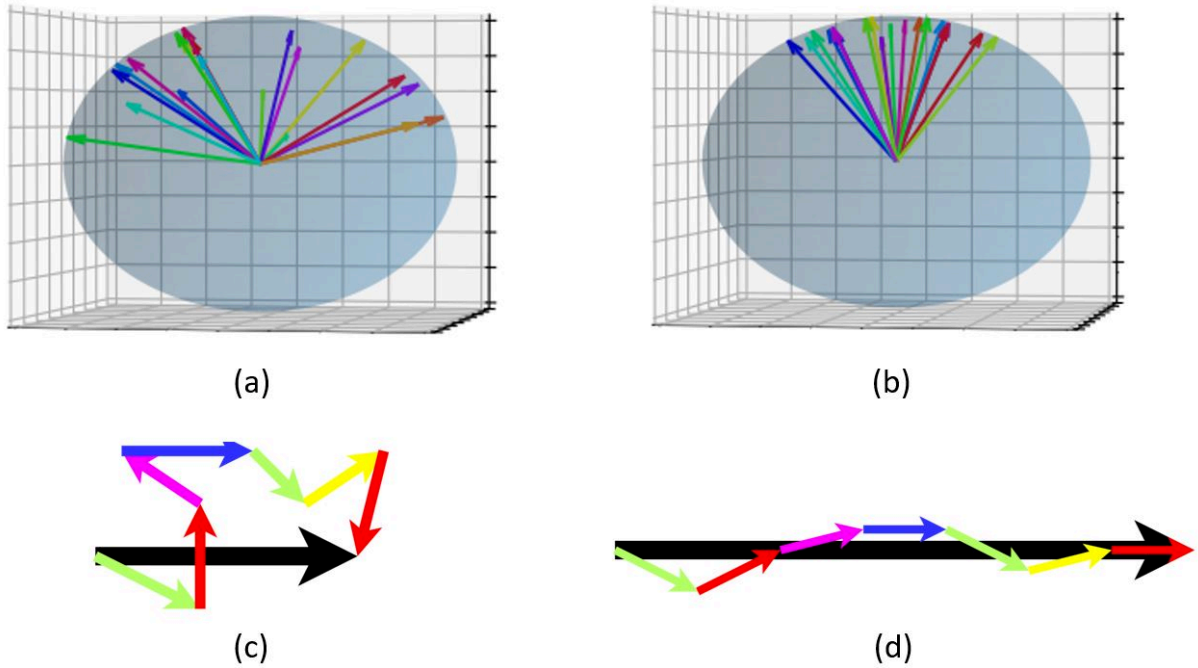


Figure 5.9: Dispersion vectors. (a) Dispersed points on a sphere. (b) polarized points on a sphere. (c) The sum of dispersed unit vectors. (d) The sum of polarized unit vectors.

the goal is to find the threshold value that discriminates the best normal and abnormal GMA. A 5-fold stratified cross-validation was used, in which the M3D values were split randomly into five subgroups (folds), ensuring that each subgroup has the same proportion of Normal and Abnormal GMA classes as the AGMA dataset (96/177). Four folds were used to select the best M3D classification threshold, and one fold was used to test the classification performance. In addition, to validate the results from 5-fold cross-validation, a Leave-One-Video-Out (LOVO) cross-validation was used. With this method, only one M3D value is set aside as a test set, while the remaining values are used for training. This process is repeated for every video in the dataset then global results and metrics are calculated as follows.

- **Specificity:** Also called the True Negative Rate (TNR), it measures the ability of a test to correctly identify negative cases called True Negatives (TN) among all the negative cases (TN + False Positives (FP)). Specificity ranges from 0 to 1, where 1 indicates perfect specificity (no false positives), and 0 indicates no specificity (all negatives are classified as positives).

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- **Sensitivity:** Also called the True Positive Rate (TPR), it measures the ability of a test to correctly identify positive cases called True Positives (TP) among all the positive cases (TP + False Negatives (FN)). Sensitivity ranges from 0 to 1, where 1 indicates perfect sensitivity (no false negatives), and 0 indicates no sensitivity (all positives are classified as negatives)

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

- **Accuracy:** It measures the correctness of a test in classifying both positive and negative cases (TP and TN). It ranges from 0 to 1, where 1 indicates perfect accuracy (all predictions are correct) and 0 indicates no accuracy (all predictions are incorrect).

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

- **F1 score:** It is especially useful when dealing with imbalanced datasets, where one class significantly outnumbers the other. It is the harmonic mean of precision ($\frac{TP}{TP+FP}$) and sensitivity (recall).

$$\text{F1} = \frac{2TP}{2TP + FN + FP}$$

- **Receiver Operating Characteristic (ROC):** It is a graphical representation for evaluating the performance of binary classification models. The ROC curve illustrates the trade-off between TPR, which is the sensitivity of the classification, and the false positive rate FPR $= 1 -$ TPR at different classification thresholds.
- **AUC:** It quantifies the overall performance of the model by measuring the area under the ROC curve. A higher AUC value indicates better model discrimination between the two classes.

In addition to these metrics, the normalized histogram of averaged 3D dispersion distributions was calculated for each population (Normal vs Abnormal). After that, the kernel density plot, which estimates the probability density function, was plotted for each histogram (see Figure 5.10). A Kolmogorov-Smirnov test was used to test the hypothesis of whether the dispersions from abnormal and normal GMs were from the same continuous distribution or not, in addition to the asymptotic p-value. Statistical analysis was performed using Python 3.7. The significance level was set at 5%.

5.5 Results and Discussion

Preterm infants with normal and abnormal GMA had a distinct M3D distribution, which was tested positively with the Kolmogorov-Smirnov test ($p < 0.001$, see Fig.5.10). The classification performance analysis of M3D for GMA revealed an AUC of 0.7840 (see Fig. 5.11). In addition, the 5-fold cross-validation method resulted in a mean M3D threshold of 0.2838 (see Table 5.2), an accuracy of 0.73 ± 0.07, and an F1 score of 0.62 ± 0.10. Moreover, when applying LOVO cross-validation, close results were obtained with an accuracy of 0.72, a sensitivity of 0.59, a specificity of 0.80, and an F1 score of 0.62. In this method, all the M3D values calculated from videos are used for training and choosing a threshold, except one test value that is compared to the calculated threshold. This repetitive comparison allowed the construction of the confusion matrix used to calculate the previous metrics. This method confirms the results obtained with the 5-fold cross-validation method by considering the abnormal GMA as the positive cases and normal GMA as the negative cases.

To verify which movement quality the M3D describes the most (complexity, variability, or fluidity), a 5-fold cross-validation was used, and results are reported in Table 5.3. The M3D had better classification performance based on the movements' complexity than their variability and fluidity with an accuracy of 0.91 and an F1 score of 0.51 due to the imbalance in data, as can be seen in Figure 5.5. In addition, when applying

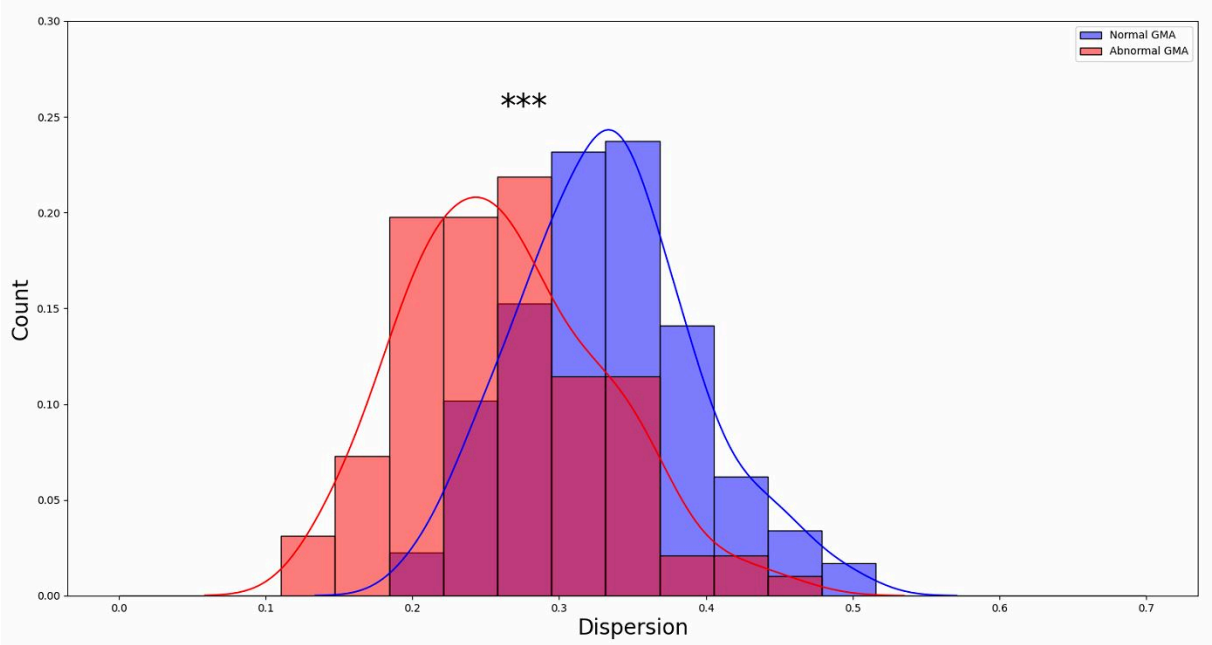


Figure 5.10: Normalized histograms and kernel density plot of the mean 3D dispersion of preterm infants with normal (n=177) *versus* abnormal GMA (n=96). ***p $<$ 0.001, with a Kolmogorov-Smirnov Test. Abbreviation: GMA, General Movement assessment.

Test fold	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean (SD)
M3D threshold	0.280	0.279	0.280	0.289	0.291	0.284 (0.006)
Accuracy	0.71	0.80	0.62	0.78	0.74	0.73 (0.07)
Sensitivity	0.60	0.70	0.47	0.65	0.62	0.61 (0.08)
Specificity	0.75	0.85	0.71	0.87	0.81	0.80 (0.07)
F1 score	0.53	0.72	0.51	0.71	0.65	0.62 (0.10)

Table 5.2: 5-fold cross-validation results and performance.

LOVO cross-validation with each movement quality, the previous results were confirmed with an F1 score of 0.47 in the classification of the movement's complexity, 0.37 in variability, and 0.35 in fluidity.

Previous studies used the MINI-RGBD dataset, which contains 12 videos of 1000 frames each, representing synthetic preterm infants' movements. As explained in section 3.5.2, the study on 3D GMA classification from Wu et al. [106] introduced an evaluation index S that discriminates normal and abnormal movements on the same principle as our method (thresholding classification). Their evaluation index was calculated based on

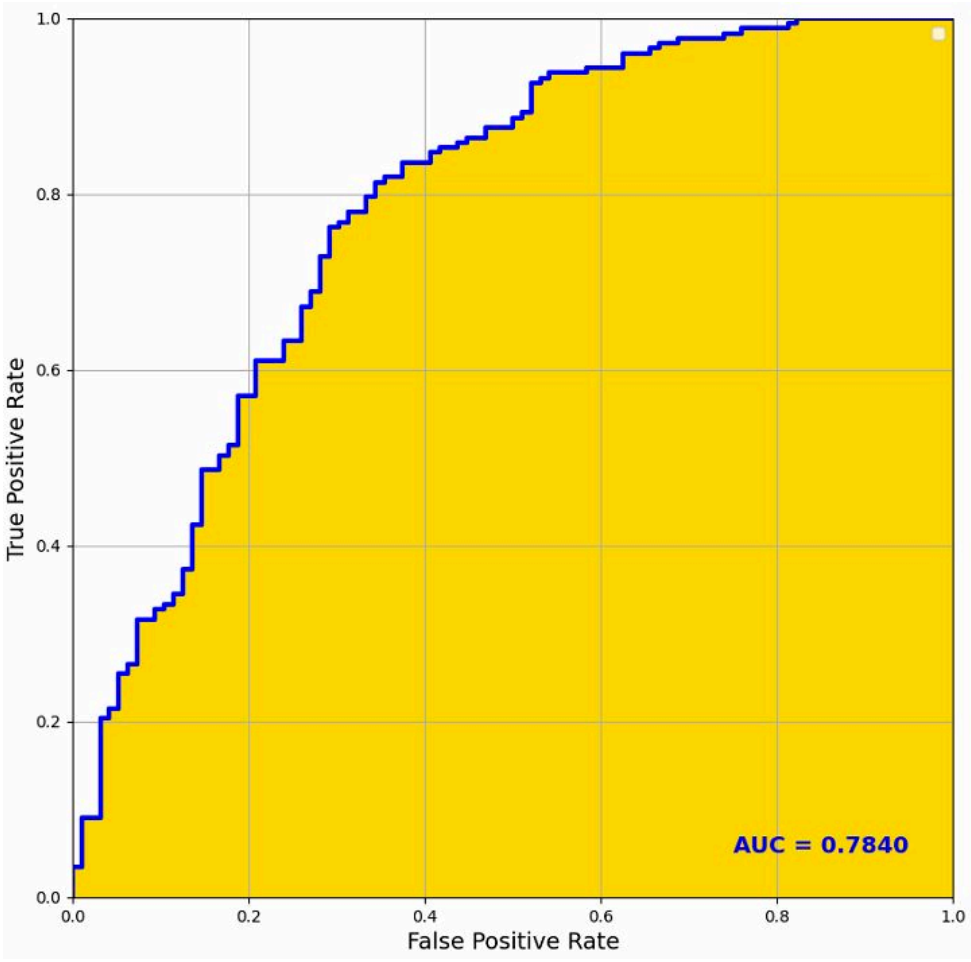


Figure 5.11: ROC curve illustrating the classification performance of the M3D for GMA. Abbreviations: AUC, area under the ROC Curve; GMA, General Movement Assessment; M3D, mean 3D dispersion.

Class	Complexity	Variability	Fluidity
M3D threshold mean	0.23	0.28	0.25
Accuracy mean (SD)	0.91 (0.03)	0.76 (0.08)	0.80 (0.04)
Sensitivity mean (SD)	0.51 (0.16)	0.77 (0.12)	0.25 (0.08)
Specificity mean (SD)	0.94 (0.03)	0.94 (0.04)	0.90 (0.02)
F1 score mean (SD)	0.51 (0.20)	0.37 (0.15)	0.28 (0.08)

Table 5.3: 5-fold cross-validation results of the classification based on movements' complexity, variability, and fluidity.

motion complexity, joint motion correlation, and small-world networks. These parameters were calculated based on using a single angle per join, which is basically the same principle as in 2D automatic GMA. Moreover, they tested and validated their method

on a very small dataset like MINI-RGBD, which makes the results relatively uncertain. The same applies to the study by McCay et al. [88], who presented neural network architectures for classifying pose-based features as normal and abnormal GMA. Both of the stated studies provided accurate results, but it was not possible to make an objective comparison due to the type and the small number of videos used. Moreover, these two studies used the Openpose network [65], which was trained on human adult images only, therefore raising the risk of keypoints detection error. Whereas for our study, in addition to the fact that it is a 3D analysis method, our pose estimation model was retrained on a manually annotated dataset of 88k real infant images.

The study with the most accurate 3D acquisition system was the one introduced by Meinecke et al. [52], who used reflecting markers with a Vicon 370 motion analysis system equipped with 6 cameras. This study included 22 infants (a relatively small dataset) and achieved a spatial precision of 2 mm. They extracted 8 parameters describing the quality of movements and reported an accuracy of 73% in classifying infants' movements, which is similar to ours (73% with one parameter). The comparison shows that the quality of the parameters is more important than their number since the M3D index outperformed the 8 parameters, including (Skewness, Cross-correlation, etc.). Moreover, the 3D mean dispersion parameter was validated on a dataset of 237 videos of infants with homogeneous gestational age and classified by an expert group composed of experienced General Movements Trust-certified assessors. Also, the spherical representation used had a major impact on the ability to compare infants' movements with different weights and ages since we normalized on the actual size of the infants' limbs. Hence, we were able to avoid biased quantification for infants with larger limbs and, consequently, larger movement amplitudes. This allowed for a more accurate and reliable comparison of the two populations. In addition, its usage can be extended to methods where the absolute distance between joints cannot be measured, such as pixels that depend on several factors like camera resolution and camera–subject distance, making the measurements not constant outside the single

video framework and the comparison of data among different subjects not valid [110]. Yet one limitation of this study is that the M3D parameter can describe the spatial characteristics and the complexity of infant movements as confirmed by the F1 scores in Table 5.3 but not their fluidity and variability, which are important parameters for GMA. The perspective of this work is to develop other parameters that can describe the fluidity and the variability of the general movements by analyzing the temporal characteristics such as velocities and acceleration vectors. These other parameters would allow a better classification of normal and abnormal general movements.

5.6 Conclusion

In this chapter, we have introduced new methods to automatically select GM sequences from long videos, based on frame differencing and pose estimation, making the GMA easier in clinical practice. Using these methods, we were able to automatically select 273 one-minute videos, classified by an expert group, and gathered in the AGMA dataset, from 129 infants born <33 weeks GA. This dataset was the first of its kind in quantity and quality (real and in a clinical environment) that allows 2D and 3D analysis and automatic assessment of general movements. Using the stereoscopic 3D framework presented in this study, in addition to a new spherical representation and normalization, we introduced the Mean 3D Dispersion parameter as a classification parameter able to discriminate normal and abnormal GMA with 73% accuracy. This parameter was validated on the AGMA dataset (273 videos) as a reliable automatic assessment tool.

6

Conclusion and Perspectives

Outline

6.1 Conclusion

In this thesis, we have observed the significant risks that premature birth imposes on preterm infants' development, ranging from motor impairments to cognitive deficits and sensory issues. Among the assessment methods used in clinical practice to identify high-risk preterm infants, Prechtl's method of GMA appeared as a valuable tool with a high degree of reliability, offering the potential for early identification of neurodevelopmental impairments and subsequently performing early interventions and targeted individual monitoring.

The subjectivity, time cost, and efforts needed to practice this method pushed the research interests toward fully automatic methods to assess infants' spontaneous movements. However, each study had a relative success as seen in the review of automatic GMA methods in chapter 3.

This thesis aimed to address the different limitations of automatic GMA by multiple means. First, it introduced a new important dataset of preterm infants' videos (282 videos from 129 infants) born before 33 weeks of gestational age (GA). These videos were reviewed and evaluated by an expert group composed of experienced General Movements Trust-certified assessors. This dataset is unique in its quality and quantity. Second, the thesis proposed the first framework for 3D infant pose estimation from stereoscopic images dedicated to preterm infants hospitalized in the Neonatal Unit. This framework is simple to implement in a clinical environment and safe for infants since it does not use any infrared cameras, unlike existing 3D pose estimation methods. It achieved a minimum mean 3D error of 1.7 cm with a 3DPCK at 5 cm of 97% based on the retrained DarkPose32 2D pose estimation network. Also, the way it predicts infants' 3D poses (2D pose + triangulation) makes it possible to use advanced deep neural networks in the future for 2D pose estimation and then 3D pose reconstruction.

On the movements classification part, this study introduced the 3D mean dispersion parameter that was able to discriminate normal and abnormal GMA with 73% accu-

racy. This parameter was validated on the AGMA dataset (282 videos) as a reliable automatic assessment tool outperforming other methods by its performance and, more importantly, its interpretability as an evaluation score. It was achieved thanks to the spherical representation used, which had a major impact on the ability to compare infants' movements with different weights and ages. This allowed for a more accurate and reliable comparison of the infants' spontaneous movements.

6.2 Perspectives

The perspectives of this work are many. First, the AGMA dataset of videos that were collected over almost 3 years is a valuable resource for the research community. However, it can not be made publicly available due to privacy concerns. Thus, one goal is to develop a method to anonymize the data while keeping its real characteristics. This will contribute to solving one of the major problems in the automatic GMA field of study, which is the significant lack of data.

Also, in the future, it will be more significant to estimate infants' 3D poses from monocular 2D images since 2D acquisition systems are more popular in the daily clinical protocol, and the dataset collected till today can serve as significant training data for 3D pose estimation models from monocular images. This will allow to democratize the GMA practice even outside the clinical usage, offering the possibility to parents to use this method using their phone cameras.

Moreover, as seen previously, the M3D parameter describes spatial characteristics and the complexity of infant movements but not their fluidity and variability, which are important parameters for GMA. So, another perspective is to develop other parameters that can describe the fluidity and the variability of the general movements by analyzing the temporal characteristics such as velocities and acceleration vectors. In fact, the spherical and normalized representation introduced in this study can be generalized on speed and acceleration vectors, and the Mean 3D Dispersion parameter could be used

as a classification threshold to discriminate movements based on their variability and fluidity. In addition, other methods can also be investigated, such as Graph Convolutional Neural Networks (GCNs) [119], where the infants' joints can be represented as nodes in a graph, which can model a wide range of relationships between the infants' limbs and characterize their overall movements.

The last perspective of this work is to test the reliability of the automatic classification not only with respect to the ground truth GMA but also with respect to their developmental outcome at 2 and 5 years of age since even if GMA has a strong predictive value, it remains a subjective assessment, and the infants' developmental outcome remains more accurate as ground truth for automatic classification.

A

Appendix

Outline

A.1 Parental Consent Form

ÉLABORATION D'UN NOUVEL OUTIL PRONOSTIQUE DEVELOPPEMENTAL DE L'ENFANT PREMATURE BASE SUR L'ANALYSE AUTOMATISEE DE LA MOTRICITE SPONTANEE

AGMA

Étude monocentrique
Version n° 1.1 du 22/01/2021

NOTICE D'INFORMATION ET DE CONSENTEMENT DU REPRESENTANT LEGAL

Coordonnées du promoteur (responsable du traitement des données)	**Coordonnées du Délégué à la Protection des Données**	**Coordonnées de l'investigateur coordonnateur**
Mme Juliette ANDRES CHU de Saint-Étienne Direction des Affaires Médicales et de la Recherche Bâtiment 31 – Hôpital Bellevue 42055 SAINT-ÉTIENNE Cedex 2 Tél : 04 77 12 70 50	Direction du système d'information Rue Bossuet 42055 Saint-Etienne Cedex 2 rgpd-dpd@chu-st-etienne.fr	Docteur GIRAUD Antoine Service de Néonatologie CHU de Saint-Étienne 42055 Saint-Étienne Cedex 2 Tél : 04 77 82 89 51

Madame, Monsieur,

La prématurité est un facteur de risque de troubles du neurodéveloppement. La valeur pronostique des différents examens réalisés en période néonatale (échographie transfontanellaire, électroencéphalogramme, IRM cérébrale) est faible.

Objectif de l'étude

C'est pourquoi, nous souhaitons développer un nouvel outil pronostique, basé sur l'analyse de la motricité spontanée. Nous vous proposons d'intégrer votre enfant dans notre protocole de recherche, qui consiste à le filmer durant une heure à trois reprises au cours de son séjour dans le service de Néonatologie. L'acquisition de ses mouvements permettra de développer un outil d'analyse automatisée. Nous pensons que ce nouvel outil permettra d'améliorer l'évaluation du pronostic précoce des enfants prématurés.

Déroulement

Si vous acceptez la participation de votre enfant à l'étude, trois sessions d'acquisition auront lieu durant son séjour dans le service de Néonatologie. En pratique, votre enfant sera allongé en couche sur le dos dans un berceau chauffant et filmé à l'aide d'une caméra spécifique dans le service de Néonatologie. Aucun geste douloureux ou invasif ne sera réalisé. En cas de pleurs ou d'inconfort, votre enfant pourra être rassuré par vous ou les soignants (comme habituellement durant l'hospitalisation) et l'enregistrement sera suspendu.
La participation à cette étude n'aura aucune conséquence sur la durée de l'hospitalisation. De plus, les données cliniques pertinentes caractérisant votre enfant et le déroulement de la grossesse seront recueillies de manière anonyme à partir du dossier médical de votre enfant.

Les résultats complets de l'analyse réalisée en routine appelée « méthode de Prechtl », à laquelle nous nous référencerons durant cette étude, vous seront remis lors de la consultation systématique à un mois de la sortie d'hospitalisation.

Bénéfices et risques.

Cette étude ne présente aucun risque pour votre enfant. Elle ne lui apportera aucun bénéfice.

Informations réglementaires.

Vous disposez d'un délai de réflexion de trois jours avant de donner votre réponse quant à la participation de votre enfant à cette recherche.

Dans le cadre de cette recherche, un traitement informatique des données personnelles de votre enfant (incluant notamment les caractéristiques périnatales et de l'accouchement) va être mis en œuvre afin de pouvoir répondre aux objectifs scientifiques de cette recherche, dans une finalité d'intérêt public. Dans ce but, les données médicales le concernant seront transmises au promoteur de la recherche ou aux personnes agissant pour son compte, en France. Ces données seront identifiées par un numéro ainsi qu'un code fait de la 1ère lettre de son nom et de son prénom. Ces données pourront également, dans des conditions expérimentales assurant leur confidentialité, être transmises aux autorités de santé françaises, à d'autres services du CHU de Saint-Étienne, et à des laboratoires de recherche de l'université de Saint-Étienne.

Conformément au Règlement Européen n°2016/679 sur la Protection des Données, vous pouvez :

- demander à avoir accès, à rectifier, à recevoir sous un format lisible numériquement ou à effacer les données concernant votre enfant
- vous opposer au recueil et à la transmission des données de votre enfant ou limiter l'utilisation de ces données uniquement à cette étude ou à d'autres situations précises
- en cas de désaccord, procéder à une réclamation auprès de la Commission National de de l'Informatique et des Libertés, 3 Place de Fontenoy - TSA 80715 - 75334 PARIS ou sur https://www.cnil.fr/webform/adresser-une-plainte

Les données de votre enfant (incluant notamment les caractéristiques périnatales et de l'accouchement) seront conservées jusqu'à la rédaction du rapport final de la recherche. Conformément à l'arrêté du 11 août 2008 fixant la durée de conservation par le promoteur et l'investigateur des documents et données relatifs à une recherche biomédicale autre que celle portant sur des médicaments à usage humain, elles seront ensuite archivées durant au moins 15 ans.

Votre enfant pourra participer simultanément à une autre étude. Il ne sera pas nécessaire qu'il respecte un délai à la fin de cette étude avant de pouvoir participer à une autre recherche.

Vous pouvez également accéder directement ou par l'intermédiaire d'un médecin de votre choix à l'ensemble des données médicales de votre enfant en application des dispositions de l'article L1111-7 du Code de la Santé Publique. Ces droits s'exercent auprès du médecin qui suit votre enfant dans le cadre de la recherche et qui connaît votre identité.

Vous êtes libre de refuser ou d'interrompre la participation de votre enfant à cette étude à tout moment sans encourir aucune responsabilité ni aucun préjudice de ce fait et sans avoir à vous justifier. Cela n'altèrera pas la qualité des soins qui seront prodigués à votre enfant et ne modifiera pas vos relations avec l'ensemble de l'équipe soignante. En cas d'interruption de l'étude, les informations concernant votre enfant seront conservées sauf opposition de votre part (et dans ce cas, elles ne pourront être supprimées qu'à condition que cela ne compromette pas gravement la réalisation des objectifs de la recherche).
Cette étude a reçu l'accord du Comité de Protection des Personnes Sud-Est II le 10/02/2021.

Vous remerciant par avance de la confiance que vous nous témoignez, nous restons à votre disposition au 04 77 82 89 51 pour tout renseignement complémentaire concernant cette étude.

Docteur GIRAUD Antoine

Je soussigné(e) M./M^{me} (Nom, Prénom) .....
déclare avoir été bien informé(e) sur l'étude "Élaboration d'un nouvel outil pronostique développemental de l'enfant prématuré basé sur l'analyse automatisée de la motricité spontanée - AGMA ». J'accepte que mon enfant participe à cette étude dans les conditions précisées ci-dessus.

Fait à ……., en deux exemplaires dont un est remis à l'intéressé(e).

	Médecin	Père	Mère
Date	……/…../……	……/…../……	……/…../……
Nom			
Prénom			
Signature			

AUTORISATION DE PRISE DE VUE

Sujet :.....

AUTORISATION CHU
AUTORISATION DE PRISES DE VUE ET D'INTERVIEWS

Vu l'article 9 du Code civil relatif au respect de la vie privée ;
Vu l'article L.1110-4 du Code de la santé publique relatif au respect de la vie privée et au secret des informations concernant le patient ;
Vu l'article L.1112-1 du Code de la santé publique relatif au secret médical et à la protection de la confidentialité des informations ;
Vu l'article R1112-47 du Code de la santé publique relatif à l'accès des journalistes et visiteurs aux malades ;
Vu la Circulaire DHOS du 2 mars 2006 relative aux droits des personnes hospitalisées et comportant une charte de la personne hospitalisée (notamment dans son article 9) ;

Je soussigné(e) : …… né(e) le……

☐ patient ☐ personnel ☑ représentant légal

demeurant :……

Adresse mail …….
autorise : le Centre Hospitalier Universitaire de Saint-Etienne

à : ☑ filmer ; ☐ photographier ; ☐ interviewer ;

et à utiliser l'image de :
☐ ma personne
☑ l'enfant …… ……..…...dont je suis le représentant légal
☐ le majeur protégé ……..... ……...dont je suis le représentant légal

En outre, dans le cadre de la communication générale de l'établissement, j'autorise, gracieusement et sans contrepartie, le CHU de Saint-Étienne à utiliser et diffuser les prises de vues photographiques / filmées qui ont été prises, dans* :

Le journal interne
Les sites internet et intranet
Les réseaux sociaux (Facebook, Twitter…etc.)
Les documents institutionnels remis aux patients et usagers

Les documents pédagogiques diffusés lors de congrès, cours, colloques, conférence (

La presse …….. dans le cadre exclusif du sujet exposé ci-avant
Autre……..

Fait à Saint- Etienne le……,

Signature précédée de la mention « lu et approuvé »

Merci **retourner** ce document dûment complété et **signé** à
CHU de Saint-Etienne, Direction Générale –Service Communication Hôpital Bellevue

A.2 AGMA Study Description

CHU PROMOTEUR
Protocole n° 20CH234
N°IDRCB : 2020-A03335-34

ÉLABORATION D'UN OUTIL PRONOSTIQUE DEVELOPPEMENTAL DE L'ENFANT PREMATURE BASE SUR L'ANALYSE AUTOMATISEE DE LA MOTRICITE SPONTANEE

AGMA

Étude monocentrique

VERSION N°1.1 du 22/01/2021

Investigateur Principal :

Docteur GIRAUD Antoine
Service de Néonatologie
Équipe d'Accueil SNA-EPIS 4607
CHU de Saint-Étienne
42055 SAINT-ÉTIENNE Cedex 2
Tel : 04 77 82 94 95

Co-investigateurs :

Professeur PATURAL Hugues, Service de Néonatologie – Équipe d'Accueil SNA-EPIS 4607.

Docteur CHABRIER Stéphane, Service de Médecine Physique et de Réadaptation Pédiatrique – Laboratoire U1059 INSERM - SAINBIOSE.

Expérimentateurs :

Professeur ALATA Olivier, Laboratoire UMR CNRS 5516 - Hubert Curien.

Professeur GAUTHERON Vincent, Service de Médecine Physique et de Réadaptation Pédiatrique – Laboratoire EA 7424 - LIBM.

SOUALMI Ameur, Équipe d'Accueil SNA-EPIS 4607.

Chef de projet URCIP :

F.PETIT

Bâtiment Recherche – Hôpital Nord
42055 Saint-Étienne Cedex 2
Tel : 04 77 12 08 26 (secrétariat)

Promoteur :

CHU de Saint-Étienne
42055 SAINT-ÉTIENNE Cedex 2.

Représentant du promoteur (responsable du traitement des données) :

Mme Juliette ANDRES
Direction des Affaires Médicales et de la Recherche
Pavillon 31 – Hôpital de Bellevue
CHU de Saint-Étienne
42055 SAINT-ÉTIENNE Cedex 2.

CPP SUD-EST II

Groupement Hospitalier Est
Bâtiment Pinel
59 Boulevard Pinel
69500 BRON

PAGE DE SIGNATURE DU PROTOCOLE

Titre de l'étude : Élaboration d'un outil pronostique développemental de l'enfant prématuré basé sur l'analyse automatisée de la motricité spontanée – AGMA.

Version du protocole et date : 1.1 du 22/01/2021

N° interne : 20CH234

Ce protocole a été lu et approuvé à la date notée ci-dessous
Les parties s'engagent à mener la recherche conformément au protocole et aux dispositions législatives et réglementaires en vigueur.

Représentant du promoteur :
Mme Juliette ANDRES
Direction des Affaires Médicales et de la Recherche
Bât 31 - Hôpital Bellevue
CHU de Saint-Étienne
42055 SAINT-ÉTIENNE Cedex 2

Date : …./…../202…
Signature
P/ Le Directeur Général et par délégation,
Le Directeur Adjoint des Affaires Médicales
Et de la Recherche,

Investigateur principal
Dr GIRAUD Antoine
Service de Néonatologie
Hôpital Nord
CHU de Saint-Étienne
42055 SAINT-ÉTIENNE Cedex 2

Date : 22/01/2021
Signature

SOMMAIRE

I - RATIONNEL

La prématurité, définie comme une naissance avant 37 semaines d'aménorrhée (SA), concerne 6% des naissances vivantes en France (1) et 10% dans le monde (2).

Elle représente un facteur de risque majeur de complications neurodéveloppementales, telles que la paralysie cérébrale, un retard cognitif, des troubles du langage, et des troubles du comportement (3,4). La prévalence de la paralysie cérébrale à 2 ans d'âge corrigé chez les enfants prématurés nés avant 32 SA est supérieure à 5% (4), soit un risque 50 fois supérieur à celui des enfants nés à terme (3). Cependant, les troubles globaux du neurodéveloppement hors paralysie cérébrale représentent actuellement la complication neurologique la plus fréquente de la prématurité, ceux-ci concernant plus de 40% des enfants nés avant 32 SA à 2 ans d'âge corrigé (4).

Les évaluations cliniques et neuroradiologiques proposées durant la période néonatale pour apprécier le pronostic développemental des enfants nés prématurément sont peu sensibles et peu spécifiques (5). L'incertitude sur le devenir développemental de cette population rend difficile la mise en place d'un suivi individualisé et induit une angoisse parentale durable qui influe en retour négativement sur les liens d'attachement et le développement de l'enfant (6).

L'analyse de l'activité motrice spontanée selon la méthode de Prechtl est actuellement l'évaluation la plus prédictive du développement de l'enfant prématuré. Elle présente une sensibilité de prédiction de la paralysie cérébrale de 80% à 100% (7) et constitue également un marqueur prédictif du développement global ultérieur (8). Cette méthode est basée sur l'analyse qualitative de la motricité spontanée par un comité d'experts lors de sessions de visionnage de séquences vidéo. Lors de ces sessions, la motricité spontanée est évaluée de manière subjective à l'aide de trois critères que sont la complexité, la variabilité, et la fluidité du mouvement. Ainsi, la motricité spontanée est jugée normale si les mouvements sont suffisamment complexes, variables et fluides, en fonction de l'âge de l'enfant en SA (Annexe 4). Celle-ci est jugée anormale si l'enfant présente un répertoire moteur pauvre – c'est-à-dire des mouvements insuffisamment complexes, variables et fluides – ou s'il produit des schémas moteurs anormaux – mouvements chaotiques, co-contractions synchrones des membres – (7,8). Néanmoins, les séquences de motricité spontanée n'étant pas présentent de manière continue, il est recommandé de filmer l'enfant durant une heure pour pouvoir extraire deux à trois séquences d'intérêt d'une durée d'une à deux minutes chacune (7). C'est pourquoi aucune équipe française de néonatologie n'effectue actuellement d'analyse de la motricité spontanée en routine clinique. Actuellement dans le service de de Néonatologie du CHU de Saint Étienne, seuls quelques enfants à très haut risque neurodéveloppemental peuvent bénéficier de cette analyse de la motricité spontanée. Celle-ci s'effectue après information et recueil de l'autorisation de prise de vue des titulaires de l'autorité parentale (Annexe 5).

Le but de notre étude est de concevoir un outil pronostique précoce et novateur du développement de l'enfant prématuré s'appuyant sur l'analyse quantitative de la motricité spontanée, basée sur la méthode de Prechtl, méthode présentant actuellement la meilleure valeur pronostique du développement ultérieur. Il n'est pas prévu d'effectuer de suivi des enfants dans le cadre de cette étude, la valeur pronostique de l'analyse de la motricité spontanée étant déjà établie (7,8).

L'objectif principal est de quantifier de manière automatisée la complexité (variations spatiales), la variabilité (variations temporelles), et la fluidité (distribution uniforme) du mouvement en fonction de l'âge des enfants nés avant 33 SA.

Le plan expérimental est le suivant. Tout enfant né prématurément avant 33 SA et hospitalisé dans le service de Néonatologie du CHU de Saint-Étienne sera retenu en vue d'une éventuelle inclusion dans l'étude. Les enfants inclus bénéficieront de trois séquences d'enregistrement réparties durant leur séjour dans le service de Néonatologie. Les enfants nés avant 33 SA sont les plus à risque de survenue de complications neurodéveloppementales (4) et sont suivis de manière standardisée jusqu'à l'âge de sept ans dans le cadre du Réseau SEVE. Les évaluer uniquement lors de leur séjour dans le service de Néonatologie permet d'inclure uniquement des enfants stables et sans support ventilatoire. Schématiquement, les enfants prématurés restent hospitalisés jusqu'au terme (40 SA) ; dans les services de Réanimation Néonatale ou de Soins Intensifs de Néonatologie lorsqu'ils ont besoin d'un support ventilatoire, puis dans le service de Néonatologie lorsqu'ils ne nécessitent plus de support ventilatoire. Ce temps d'hospitalisation prolongé et prévisible offre l'opportunité de présenter puis de réaliser l'étude dans des conditions satisfaisantes.

L'hospitalisation dans le service de Néonatologie a lieu à des âges très différents selon les enfants, et la sortie d'hospitalisation a lieu sensiblement au même âge – autour du terme (40 SA). L'espace entre les séquences sera donc variable entre les enfants inclus. Le fait de définir ces séquences d'enregistrement en fonction des dates d'hospitalisation assurera une répartition homogène des différents enregistrements en fonction de l'âge et n'aura pas d'incidence sur l'interprétation des résultats dans la mesure où ceux-ci seront exprimés en fonction de l'âge exprimé en SA.

Le critère d'évaluation principal sera composite et regroupera trois conditions, mesurées à trois reprises :

- **La complexité du mouvement sera définie par la valeur de variation spatiale, correspondant à la proportion de visites sur l'espace que peuvent occuper les membres de l'enfant. Il s'agit d'un rapport de deux volumes et donc, cette mesure sera sans unité. Elle aura une valeur entre 0 et 1.**

- **La variabilité du mouvement sera définie par l'écart-type des intensités des vitesses des membres. Nous calibrerons notre système de manière à pouvoir faire une mesure en $m.s^{-1}$.**

- **La fluidité du mouvement sera évaluée à partir de la forme des distributions des intensités et des directions des mouvements. Nous produirons une valeur normalisée entre 0 et 1.**

De tels abaques n'existent pas à l'heure actuelle, et sont nécessaires pour caractériser les mouvements spontanés de chaque enfant. Les critères d'évaluation secondaires sont la détection et la quantification des schémas moteurs anormaux – associés à des troubles du neurodéveloppement (7) – et la détermination de seuils pathologiques à partir de la distribution des valeurs normalisées de complexité, de variabilité, et de fluidité du mouvement.

Ces critères d'évaluation seront mesurés lors de trois séquences d'enregistrement distinctes lors du séjour dans le service de Néonatologie. Chaque enregistrement durera une heure, soit la durée recommandée dans la cadre de l'évaluation de la motricité spontanée permettant d'acquérir trois séquences motrices de qualité (7). Le signal sera obtenu par caméra stéréo 2k ZED 2 (Stereolabs, USA). Il s'agit d'une caméra deux vues décalées de manière concomitante, ce qui permet d'obtenir deux images en couleur de la scène ainsi qu'une image de profondeur calculée par stéréovision. Cette information de profondeur permet d'obtenir une localisation en trois dimensions de l'enfant.

L'enregistrement aura lieu dans une salle dédiée du service de Néonatologie, afin de ne pas rompre la surveillance continue dont bénéficie ces enfants. Le traitement de ces séquences vidéo sera effectué par Ameur Soualmi, doctorant à temps plein sur ce projet de recherche.

Les perspectives de ce travail, au-delà d'un dépôt de brevet et des publications dans des revues internationales attendus pour un tel projet, sont l'amélioration de l'évaluation du pronostic développemental des enfants prématurés et l'instauration d'un suivi individualisé de ces enfants.

II - OBJECTIFS

2.1 Objectif principal

Quantifier de manière automatisée la complexité (variations spatiales), la variabilité (variations temporelles), et la fluidité (distribution uniforme) du mouvement en fonction de l'âge en SA.

2.2 Objectif secondaire

Comparer notre analyse quantitative à l'analyse qualitative selon la méthode de référence (Pretchl).

III - CONCEPTION DE LA RECHERCHE

3.1 Critères d'évaluation

3.1.1 Critère d'évaluation principal

Critère composite regroupant trois conditions, mesurées à trois reprises :

- La complexité du mouvement sera définie par la valeur de variation spatiale, correspondant à la proportion de visites sur l'espace que peuvent occuper les membres de l'enfant. Il s'agit d'un rapport de deux volumes et donc, cette mesure sera sans unité. Elle aura une valeur entre 0 et 1.

- La variabilité du mouvement sera définie par l'écart-type des intensités des vitesses des membres. Nous calibrerons notre système de manière à pouvoir faire une mesure en $m.s^{-1}$.

- La fluidité du mouvement sera évaluée à partir de la forme des distributions des intensités et des directions des mouvements. Nous produirons une valeur normalisée entre 0 et 1.

3.1.2 Critère d'évaluation secondaire

Critère composite regroupant quatre conditions, mesurées à trois reprises :

- Erreur globale de classification
- Précision
- Rappel
- F1-score

3.2 Méthodologie de la recherche

Étude monocentrique, prospective, observationnelle, chez les enfants nés avant 33 SA hospitalisés dans le service de Néonatologie du CHU de Saint-Étienne.

3.3 Mesures prises pour réduire et éviter les biais

L'ensemble des enfants nés avant 33 SA hospitalisés dans le service de Néonatologie ne présentant pas de critère de non-inclusion seront proposés à l'étude, de manière consécutive. La population des enfants nés avant 33 SA au CHU de Saint Étienne est considérée comme représentative des enfants prématurés français (4,9).
L'acquisition des mouvements et le traitement du signal sera identique pour tous les enfants.

Les objectifs de cette étude sont descriptifs et les analyses sont réalisées selon l'âge en SA, comme décrit dans la méthode de référence. Aucune comparaison en sous-groupe n'est prévue dans cette étude. Le recueil des caractéristiques telles que le poids de naissance, la présence d'une chorioamniotite ou d'un syndrome polymalformatif est réalisé afin de décrire au mieux notre population. Il n'y aura donc pas de facteur confusion à prendre en compte dans notre analyse.

L'analyse qualitative de la motricité spontanée est réalisée par un comité d'experts. Cette analyse est réalisée en aveugle de toutes les données cliniques en-dehors de l'âge en SA lors de l'enregistrement. Ce comité d'experts regroupe une dizaine de professionnels de santé (pédiatres, kinésithérapeutes, psychomotriciens) du bassin stéphanois ayant un certificat de formation à l'évaluation des Mouvements Généraux selon la méthode de Prechtl délivré par le GM Trust (http://general-movements-trust.info).

3.4 Le déroulement

3.4.1 Déroulement pour un patient :

La sélection :
Tout enfant né avant 33 SA hospitalisé dans le service de Néonatologie du CHU de Saint-Étienne sera retenu en vue d'une éventuelle inclusion dans l'étude.

L'inclusion :
La procédure d'inclusion aura lieu lors de l'hospitalisation dans le service de Néonatologie du CHU de Saint-Étienne. La vérification des critères d'inclusion et de non- inclusion, l'explication de l'étude, la distribution de la notice d'information et le recueil du consentement des titulaires de l'autorité parentale seront effectuées par l'investigateur.

Le suivi :
L'ensemble de l'étude se déroulera durant l'hospitalisation dans le service de Néonatologie du CHU de Saint-Étienne. Aucune consultation supplémentaire ne sera réalisée pour cette étude. Aucun examen invasif ou irradiant supplémentaire ne sera réalisé. La durée d'hospitalisation des enfants ne sera pas impactée par ce projet.

Trois séquences d'enregistrement d'une heure seront effectuées durant l'hospitalisation, dans une salle dédiée à l'intérieur du service de Néonatologie. La première se déroulera quelques jours après l'hospitalisation, la deuxième en milieu de séjour, et la troisième quelques jours avant la sortie d'hospitalisation. L'hospitalisation dans le service de Néonatologie a lieu à des âges très différents selon les enfants, et la sortie d'hospitalisation a lieu sensiblement au même âge – autour du terme théorique. L'espace entre les séquences sera donc variable entre les enfants inclus.

Durant celles-ci, l'enfant vêtu d'une couche sera placé dans un berceau chauffant en décubitus dorsal, ce qui est habituellement le cas durant l'hospitalisation. L'ECG sera monitoré en temps réel à l'aide de trois électrodes, de la même manière que durant le reste de l'hospitalisation. La saturation en oxygène pourra être monitorée durant l'enregistrement en cas d'indication médicale, de la même manière que durant l'hospitalisation. Au final, que l'enfant soit inclus ou non, il bénéficiera de la même surveillance indiquée par son état clinique.

L'activité motrice sera acquise à l'aide de la caméra stéréo 2K ZED 2 (Stereolabs, USA), reliée à un ordinateur portable. Un microphone sera utilisé pour détecter d'éventuels pleurs durant l'acquisition. Toute stimulation visuelle ou auditive sera évitée durant l'enregistrement, afin de laisser s'exprimer la motricité spontanée de l'enfant. Les parents pourront être présents lors de l'acquisition. En cas de pleurs ou d'inconfort, l'enfant sera rassuré par les parents ou les soignants – comme habituellement durant l'hospitalisation – et l'acquisition suspendue.

Cette méthode sera donc strictement indolore et non-invasive, dans la mesure où les enfants seront allongés dans un berceau chauffant dans le service de Néonatologie lors de l'acquisition des séquences et ne seront soumis à aucune stimulation supplémentaire.

Les résultats complets de l'analyse qualitative de l'activité motrice spontanée selon la méthode de Prechtl seront remis aux parents lors de la consultation systématique à un mois de la sortie d'hospitalisation, à l'occasion d'une consultation systématique au CHU de Saint Étienne dont bénéficient tous les enfants nés avant 33 SA grâce au réseau de suivi SEVE.

Le traitement de ces séquences vidéo sera effectué par Ameur Soualmi, doctorant à temps plein sur ce projet de recherche.

La fin d'étude :
L'étude se terminera à la fin de la troisième séance d'enregistrement, soit avant la sortie du patient du CHU de Saint-Étienne.

3.4.2 Durée de participation à l'étude pour un patient :

La durée de l'étude pour un patient dépendra de l'enchainement des trois séquences d'enregistrement, mais sera de toute façon intégrée à son hospitalisation. Au maximum, l'étude durera 3 mois pour un patient.

3.4.3 Participation à une autre étude simultanément ou à la fin de celle-ci :

Le patient pourra participer simultanément à une autre étude. Il ne sera pas nécessaire qu'il respecte un délai à la fin de cette étude avant de pouvoir participer à une autre recherche.

3.5 Description des règles d'arrêt définitif ou temporaire

Si les inclusions n'ont pas commencé dans les 2 ans qui suivent l'obtention de l'autorisation du CPP, l'étude sera arrêtée. Si le taux d'inclusion est insuffisant et non justifié, le promoteur pourra décider de l'arrêt de l'étude.

IV - SÉLECTION ET EXCLUSION DES PERSONNES DE LA RECHERCHE

4.1 Critères d'inclusion

- Naissance avant 33 SA ;
- Hospitalisation dans le service de Néonatologie du CHU de Saint-Étienne ;
- Patient affilié ou ayant droit d'un régime de sécurité sociale ;
- Autorisation de prise de vue des titulaires de l'autorité parentale (Annexe 5) ;
- Consentement de participation à l'étude des titulaires de l'autorité parentale.

4.2 Critères de non-inclusion des personnes qui se prêtent à la recherche.

- Nécessité d'un support ventilatoire ;
- Difficulté de régulation thermique contre-indiquant l'utilisation d'un berceau chauffant ;
- Refus d'autorisation de prise de vue des titulaires de l'autorité parentale (Annexe 5) ;
- Refus de participation à l'étude des titulaires de l'autorité parentale.

4.3 Justification de la participation de personne incapable de donner leur consentement éclairé ou d'autres populations spéciales et la description

- Cette étude s'intéresse exclusivement aux enfants.

4.4 Détailler les critères de sortie d'essai (arrêt de l'essai) et les procédures spécifiques :

Si un titulaire de l'autorité parentale suspend la participation de son enfant à l'étude, la date et la cause de sortie de l'étude seront notés dans le cahier de recueil des données. Aucune procédure spécifique supplémentaire ne sera nécessaire.

V - ANALYSE STATISTIQUE

5.1. Nombre de sujets nécessaires

Dans le cadre de cette étude descriptive non comparative, plus le nombre de sujets inclus sera important, plus les analyses seront précises. Compte-tenu de la durée de l'étude – 3 ans – et du nombre de naissances prématurées avant 33 SA au CHU de Saint Étienne – 120 par an –, notre objectif est d'obtenir 900 points pour chaque condition durant l'étude, correspondant à 3 enregistrements chez 300 enfants.
Compte-tenu de la relation établie entre l'équipe soignante et la famille durant l'hospitalisation et le caractère non-invasif de ce protocole de recherche, un ratio d'enfants inclus/incluables autour de 80% est réaliste.

5.2.Description des méthodes statistiques

5.2.1. Critère de jugement principal

A partir de ces images obtenues par stéréovision, des fonctions d'analyse, déjà développées pour cette caméra, offrent une représentation filaire (ou squelette) d'un corps avec, en particulier, la localisation 3D de points d'intérêt indiquant la localisation des mains, des coudes, des pieds et des genoux. Ces informations vont nous permettre de connaître l'espace visité par l'enfant, ainsi que les vecteurs vitesses (intensités et directions) en exploitant les différences de positions d'une image à l'autre dans la vidéo.

A chaque acquisition, nous aurons ainsi la possibilité d'avoir des distributions pour ces grandeurs : points des visites, intensités et direction des vecteurs vitesses pour les différents membres. A partir de ces ensembles de valeurs,

- la complexité du mouvement sera obtenue comme la proportion de visites sur l'espace que peuvent occuper les membres de l'enfant. L'analyse des localisations 3D des membres supérieurs permettra de connaître un volume défini par la surface des positions limites visitées. Nous ferons donc un rapport entre ce volume et le volume possible. Une valeur par membre peut être obtenue. Nous ferons alors une moyenne (arithmétique, harmonique, et géométrique) afin de donner une valeur finale.
- la variabilité de la vitesse du mouvement sera obtenue à l'aide de la dispersion de l'intensité des vitesses par l'estimation de la variance.
- la fluidité du mouvement sera déterminée par la forme globale de la distribution des intensités des vitesses ainsi que de celle de la distribution des directions. Pour un enfant présentant des mouvements normaux, toutes les intensités possibles et toutes les directions sont censées se produire, se rapprochant ainsi de distributions uniformes. Pour évaluer la forme des distributions des intensités, nous utiliserons le *kurtosis*[1] qui possède une valeur minimale pour une distribution uniforme et maximale pour des distributions possédant des pics. Pour évaluer la forme des distributions de direction, nous utiliserons le paramètre de concentration de la distribution de Fisher[2]. La condition sera alors obtenue par une combinaison de ces deux paramètres (moyenne arithmétique, harmonique ou géométrique).

Ces calculs, réalisés pour tous les enfants, nous permettront d'avoir des distributions des conditions pour chaque classe d'âge en SA. Le groupe d'experts qui visionnera les vidéos permettra de fournir une annotation qualitative (mouvements normaux / anormaux pour chaque condition) et donc de disposer d'une « vérité terrain ». L'objectif secondaire consiste à trouver des seuils, pour chaque semaine, sur les critères quantitatifs permettant de séparer les mouvements normaux et anormaux. A cet effet, nous étudierons l'usage des conditions indépendamment les unes des autres, ce qui permettra de déterminer des seuils pour chaque condition, ainsi que conjointement. Dans le premier cas d'étude (hypothèse d'indépendance), nous travaillerons à l'aide de distributions unidimensionnelles et dans le deuxième cas d'étude, nous travaillerons à l'aide de probabilités conjointes, c'est à dire des distributions multidimensionnelles, de manière à prendre en compte les éventuelles dépendances entre les conditions (c.à.d. complexité, variabilité et fluidité).
Les méthodes envisagées sont les méthodes de classification bayésiennes paramétriques et non paramétriques, nécessitant l'estimation d'une distribution pour chaque classe, et les machines à vecteurs de support. Cette dernière méthode sera utilisée exclusivement dans le cas multidimensionnel. Son intérêt par rapport aux méthodes bayésiennes consiste à ne pas nécessiter un modèle de distribution pour se concentrer sur la définition de la frontière entre les classes. Cela dit, nous ne détaillerons pas ici le cas multidimensionnel qui est l'extension du cas scalaire.

---

[1] Pour une variable aléatoire réelle X, $\mathrm{Kurt}[X] = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$, avec $\mu = \mathbb{E}[X]$ et $\sigma^2 = \mathbb{E}[(X-\mu)^2]$, la variance. σ est l'écart-type.

[2] Pour un vecteur aléatoire D dont l'espace de réalisation est l'ensemble des directions 3D et dont la loi de probabilité est la loi de Fisher, la densité de probabilité est définie de la manière suivante :

$$f_D(\mathbf{d}) = \frac{\kappa}{\sinh(\kappa)} e^{\kappa \boldsymbol{\mu}^T \mathbf{d}}$$

avec $\boldsymbol{\mu}$ le vecteur des directions moyennes et $\mathbf{d} \in \mathbb{R}^3$, $\|\mathbf{d}\|_2 = 1$. $\kappa > 0$ est le paramètre de concentration des directions. Plus il est faible, plus la distribution des directions est dispersée sur toute la sphère unité.

Une méthode de classification bayésienne paramétrique exploite un modèle paramétré de distribution comme, par exemple, celui de la loi de Gauss qui est paramétré par la moyenne et la variance. Nous ferons donc des tests statistiques de Kolmogorov-Smirnov sur chaque distribution de manière à choisir le modèle le plus approprié (loi de Gauss, gaussienne généralisée, distribution de Pearson).

L'approche non paramétrique permet de ne pas faire d'hypothèse de loi paramétrée. Les deux principales méthodes d'estimation non-paramétrique[3] d'une distribution sont :

- l'histogramme,
- la méthode à noyau.

A partir d'un ensemble de *N* valeurs $\{x_1, \dots, x_N\}$, normalisées entre 0 et 1 (sans perte de généralité car il suffit de remplacer 0 par le minimum des valeurs et 1 par le maximum), la formule d'estimation de la distribution à l'aide de l'histogramme, possédant des intervalles (ou classes) de largeur fixe $\{C_1, \dots, C_{N_i}\}$, N_i le nombre d'intervalles, est la suivante :

$$\hat{f}_h(x) = \frac{1}{Nh} \sum_{j=1}^{N_i} n_j \, 1_{C_j}(x)$$

avec $h = \frac{1}{m}$, la largeur des intervalles, et $n_j = \sum_{i=1}^{N} 1_{C_j}(x_i)$, l'occurrence de l'intervalle C_j. $1_{C_j}(.)$ est la fonction indicatrice de l'intervalle C_j, $j = 1, \dots, N_i$. Cette méthode nécessite de définir le nombre d'intervalles, N_i. Des études statistiques ont montré que $N_i < O\left(2\lfloor\sqrt{N} - 1\rfloor\right)$, avec $\lfloor . \rfloor$ la partie entière en valeur inférieure. Il peut aussi être obtenu en utilisant des méthodes exploitant les critères d'information. Néanmoins, bien que pratique, cette méthode présente l'inconvénient de fournir une distribution constante par morceau et donc une analyse qui discrétise les valeurs étudiées suivant les intervalles, et donc les seuils recherchés seront pris parmi ces valeurs.
La formule de la méthode à noyau est :

$$\hat{f}_K(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right)$$

avec $K(.)$ le noyau et *h* la fenêtre (ou « bandwidth » en anglais). $\hat{f}_K(.)$ possède donc les propriétés de continuité et de dérivabilité du noyau. Ainsi, si un noyau gaussien est utilisé, $\hat{f}_K(.)$ sera une estimation continue et dérivable de la distribution ce qui offre la possibilité d'obtenir des seuils sur toute la plage des valeurs envisagées. Reste alors à fixer h : il sera choisi de manière à avoir des distributions présentant un mode, c'est-à-dire de manière à éviter les modes secondaires.

Une fois que l'on dispose du modèle des distributions, il est alors possible de les utiliser pour décrire, au niveau de chaque condition, la distribution associée à la classe des mouvements normaux et la distribution associée à la classe des mouvements anormaux. La stratégie de classification bayésienne consiste à trouver le seuil permettant de minimiser un coût moyen. Différents coûts existent. Par exemple, celui qui consiste à choisir la classe par Maximum a Posteriori (MAP)[4] est associée au coût 0-1 qui consiste à dire que toutes les erreurs ont le même coût (1) et qu'un bon choix ne coûte rien (0). En faisant un choix judicieux au niveau du coût, il est possible de palier au problème de données déséquilibrées.

Dans le cas des analyses à partir de combinaisons de paramètres, nous déterminerons la frontière de décision correspondant à la classification qui est alors une fonction multidimensionnelle. Il sera alors possible de projeter cette frontière de décision dans des espaces 2D de critères combinés, ce qui donnera des courbes délimitant les régions normales et anormales. Ces courbes remplaceront les seuils des cas scalaires. Comme nous travaillons a priori dans un espace de paramètres à 3 dimensions (complexité c, variabilité v et fluidité f), la frontière de séparation entre les classes ainsi que la position d'un point de l'espace par rapport aux différentes classes peuvent être difficiles à observer, surtout si on ne dispose pas d'un écran d'ordinateur. Le fait de projeter les frontières dans des espaces 2D (c-v, c-f et v-f) permettra d'observer sur une feuille de papier la position d'un point par rapport aux classes dans les différents plans et donc d'interpréter l'information plus facilement.

[3] Nous excluons ici la méthode par les k-plus-proches-voisins car elle ne nous permettrait pas d'obtenir des valeurs de seuil.

[4] Dans le cas d'une classification binaire, la probabilité a posteriori est $P(C = c_i|x) = \frac{f_{X|c_i}(x|c_i)P(C=c_i)}{f_X(x)}$, $i = 1,2$, avec *x* la valeur du critère observée, C la variable aléatoire associée à la classe, $f_{X|c_i}(.|c_i)$ la densité associée à la classe *i*, $P(C = c_i)$ la probabilité a priori de la classe *i*. La classification par MAP consiste à choisir la classe qui possède la probabilité a posteriori la plus élevée.

5.2.2. Critère de jugement secondaire

Les différentes vidéos seront visionnées par un groupe d'experts afin de déterminer suivant la méthode de référence de Pretchl les motricités spontanées normales et anormales. Cette détection nous servira de « vérité terrain » pour définir des seuils sur les données quantitatives. A cet effet, nous utiliserons des méthodes de classification supervisées. Les conditions du critère d'évaluation secondaires seront les conditions classiques pour évaluer ces méthodes : matrice de confusion puis à partir de cette matrice, le calcul de l'erreur globale de classification, de la précision, du rappel et du F1-score, qui en est la moyenne harmonique.

Nous étudierons ainsi différentes méthodes en les confrontant aux annotations du groupe d'experts, issus de la méthode de Prechtl. Pour une méthode de classification binaire (0 pour négatif et 1 pour positif), le critère d'évaluation secondaire sera un critère composite regroupant les conditions suivantes :

- l'erreur globale de classification $e = \frac{FP+FN}{TP+FP+TN+FN}$,
- le rappel $r = \frac{TP}{TP+FN}$,
- la précision $p = \frac{TP}{TP+FP}$,
- le F1-score $f_1 = 2\frac{p \times r}{p+r}$.

Avec : TP (*True Positive* – le nombre de cas classé 1 qui appartiennent bien à la classe 1), FP (*False Positive* – le nombre de cas classé 1 qui appartiennent à la classe 0), TN (*True Negative* – le nombre de cas classé 0 qui appartiennent bien à la classe 0) et FN (*False Negative* – le nombre de cas classé 0 qui appartiennent à la classe 1).

Les différentes méthodes de classification choisies seront comparées, suivant ces critères[5], à l'aide d'une procédure de *cross-validation* de type *leave one subject out* (LOSO), permettant de sélectionner la méthode qui a le meilleur comportement sur des données inconnues et de déterminer des seuils plus robustes.

5.3. Degré de signification statistique

Les résultats seront considérés comme significatifs au seuil de 5%.

5.4. Choix des personnes à inclure dans les analyses

L'analyse sera réalisée à partir de l'ensemble des patients inclus dans l'étude et sur les données disponibles.

VI - DROIT LIÉS A LA PROTECTION DES DONNÉES ET DOCUMENTS SOURCES

6.1. Protection des données

Conformément au Règlement Européen sur la Protection des Données, les personnes participant à cette recherche seront informées, via la notice d'information et le formulaire de consentement, des droits suivants :

- de la nature et de la finalité des données recueillies dans le cadre de la recherche ainsi que du délai de conservation de ces données
- de la possibilité d'arrêter l'étude à tout moment et de la conservation, par le promoteur, des informations recueillies (sauf indication contraire de la personne concernée).
- de leurs droits d'accès, de rectification, d'opposition, de limitation, d'effacement et de portabilité des données recueillies dans le cadre de la recherche. Ces droits pourront s'exercer à tout moment de la recherche soit en effectuant une demande auprès du médecin qui suit les personnes dans le cadre de la recherche (et qui contactera le promoteur) soit en déposant une demande auprès du Délégué à la Protection des Données du promoteur

[5] Prenons un exemple simple (mais avec des données déséquilibrées tout de même) pour illustrer ces valeurs. On dispose de100 échantillons de la classe 0 vs 10 échantillons de la classe 1 et TN = 98, FP = 2, FN = 8, TP = 2. C'est plutôt mauvais mais e = 10/110 ≈ 0,091 ce qui peut paraître plutôt bien. Par contre, f_1 = 2/7 = 0,287 ce qui indique une classification plutôt pas bonne (le maximum est égal à 1) et l'intérêt d'utiliser ce critère dans le cas de données déséquilibrées.

- de la possibilité, en cas de problème/désaccord, d'effectuer une réclamation auprès de la CNIL à l'adresse suivante : https://www.cnil.fr/webform/adresser-une-plainte.

Le promoteur (par l'intermédiaire du TEC ou des investigateurs) s'engage à répondre à toute demande d'accès aux données dans un délai d'un mois maximum. Par ailleurs, seul le personnel habilité par le promoteur (investigateurs, ARC, TEC) et les représentants des autorités de santé pourront avoir accès à ces informations.

Il a été précisé dans la convention que l'investigateur permettra le monitoring par le promoteur, de possibles audits et inspections des comités d'éthique, des IRB (Institutional Review Board), des autorités compétentes, en fournissant un accès direct aux documents source.

6.2. Description des variables recueillies

Les données de l'étude seront recueillies directement dans les cahiers d'observation, au fur et à mesure des visites de l'étude. Ces données seront validées par l'investigateur qui signera (manuellement ou électroniquement) les cahiers d'observation.

Les informations recueillies dans le cadre de cette étude sont systématiquement notées dans le dossier médical des enfants prématurés et permettront de caractériser notre population. Il s'agit des informations suivantes :

- Données démographiques : terme de naissance, sexe, mensurations (poids, taille, périmètre crânien) de naissance et hebdomadaires, âge (SA) ;
- Caractéristiques périnatales : âge maternel, grossesse unique ou multiple, grossesse spontanée ou induite, pathologie associée à la naissance (vasculaire, placentaire, diabète, menace d'accouchement prématuré, chorioamniotite, anomalie du rythme cardiaque fœtal), corticothérapie anténatale (complète, incomplète, non faite), prélèvement vaginal, examen cytobactériologique des urines maternelles ;
- Déroulement de l'accouchement : voie basse ou césarienne (avant ou pendant travail, cause maternelle ou fœtale), instrumental ou non, score d'Apgar à 1 et 5 minutes, réanimation en salle de naissance (oxygène, intubation, ventilation invasive ou non invasive, adrénaline) ;
- Pathologies néonatales : syndrome malformatif, anomalie génétique, pathologies respiratoires (durée de ventilation invasive et non invasive, durée d'oxygénothérapie, nombres de doses de surfactant, date d'arrêt de tout support ventilatoire), pathologies cardiovasculaires (remplissage, amines vasopressives, monoxyde d'azote), pathologies infectieuses (infections néonatales bactériennes précoce et tardive, méningite), pathologies digestives (entérocolite ulcéro-nécrosante, jours de mise à jeun), pathologies neurologiques (hémorragie cérébrale, lésions de la substance blanche, convulsions), pathologie auditive, pathologie ophtalmologique ;
- Traitements : médicamenteux, chirurgicaux, kinésithérapie, psychomotricité, alimentation (maternelle, artificielle, mixte) ;
- Constantes lors de l'enregistrement : fréquence cardiaque, SDNN (déviation standard de l'intervalle RR), fréquence respiratoire, saturation en oxygène, température ;
- Résultats des explorations neurologiques réalisées de manière systématique : échographies transfontanellaires, électroencéphalogrammes, IRM cérébrale, examen du fond d'œil.

Les données manquantes devront être justifiées (par exemple dans un tableau des violations de protocole). Toute correction apportée dans le CRF devra être traçable (barrée, datée et paraphée par le correcteur sur un CRF « papier », *via* « l'audit trial » sur un e-CRF).

6.3. Identification des données sources

Les documents sources seront constitués par l'ensemble des informations et résultats d'examens figurant dans le dossier médical des personnes participants à cette recherche. Les informations suivantes devront figurer dans le dossier médical :

- titre de l'étude,
- date d'information et d'inclusion du patient dans l'étude (date de la remise de la note d'information et du recueil du consentement des parents),

- les différentes visites du patient dans le cadre du protocole,
- la survenue des complications (Evénements Indésirables Graves).

L'investigateur s'engage à donner un accès direct à l'ensemble de ces documents aux personnes mandatées par le promoteur.

VII - CONTRÔLE ET ASSURANCE QUALITÉ

Cette étude étant dénuée de risques, aucun contrôle qualité ne sera réalisé sur ce projet.

VIII - FAISABILITE ET CALENDRIER DE L'ÉTUDE

Environ 120 enfants nés avant 33 SA sont hospitalisés chaque année dans le service de Néonatologie du CHU de Saint-Étienne. Nous anticipons un recrutement de 300 enfants sur 3 ans.

Un doctorant de l'UJM a été recruté à temps plein sur ce projet à compter du 1er Octobre 2020. Son financement est assuré par un Contrat Doctoral de l'École Doctorale SIS pour une durée de 3 ans.

Calendrier prévisionnel :
- Durée de l'étude : 3 ans ;
- Début prévisible de début des inclusions : Janvier 2021 ;
- Date de fin des inclusions : Janvier 2024 ;
- Date de fin du suivi : Janvier 2024 ;
- Analyse et validation des résultats : Septembre 2024 ;
- Rapport final : Janvier 2025 au plus tard.

IX - ORGANISATION DE L'ÉTUDE

Investigateur coordonnateur : Docteur GIRAUD Antoine, Service de Néonatologie – Équipe d'Accueil SNA-EPIS 4607.

Centre investigateur : CHU de Saint-Étienne.

Expérimentateurs :
- Professeur PATURAL Hugues, Service de Néonatologie – Équipe d'Accueil SNA-EPIS 4607 ;
- Professeur ALATA Olivier, Laboratoire UMR CNRS 5516 - Hubert Curien ;
- Professeur GAUTHERON Vincent, Service de Médecine Physique et de Réadaptation Pédiatrique – Laboratoire EA 7424 – LIBM ;
- Docteur CHABRIER Stéphane, Service de Médecine Physique et de Réadaptation Pédiatrique – Laboratoire U1059 INSERM – SAINBIOSE ;
- SOUALMI Ameur, Équipe d'Accueil SNA-EPIS 4607.

Service concerné par la recherche : Néonatologie.

Responsables de la mesure des critères d'évaluation :
- Docteur GIRAUD Antoine, Service de Néonatologie – Équipe d'Accueil SNA-EPIS 4607 ;
- Professeur PATURAL Hugues, Service de Néonatologie – Équipe d'Accueil SNA-EPIS 4607 ;
- Professeur ALATA Olivier, Laboratoire UMR CNRS 5516 - Hubert Curien.

Personnes chargées de la saisie des données :
- SOUALMI Ameur, Équipe d'Accueil SNA-EPIS 4607 ;
- Docteur GIRAUD Antoine, Service de Néonatologie – Équipe d'Accueil SNA-EPIS 4607.

Responsable de l'analyse statistique : Professeur ALATA Olivier.

X - CONSIDERATIONS ÉTHIQUES

10.1. Principes généraux

Le protocole est en conformité avec les principes d'éthique établis par la 18ème Assemblée Médicale Mondiale (Helsinki 1964) et par les amendements établis lors des 29ème (Tokyo 1975), 35ème (Venise 1983), 41ème (Hong Kong 1989), 48ème (Somerset West 1996), 52ème (Edinburg 2000), 53ème (Washington 2002), 55ème (Tokyo), 59ème (Seoul) et révisée lors de la 64ème Assemblée Médicale Mondiale (Fortaleza, Brésil, Octobre 2013). Il sera conduit conformément aux recommandations ICH de Bonnes Pratiques Cliniques.

10.2. Protection des personnes

Il s'agit d'une étude de catégorie 3 de la loi Jardé car les actes pratiqués ne sont pas invasifs et sont dénués de risque. Le protocole sera soumis à l'avis d'un Comité de Protection des Personnes (CPP) avant de débuter les inclusions.

Les titulaires de l'autorité parentale se verront remettre une notice d'information (annexe 3) leur indiquant leurs droits quant aux données collectées.

S'il est prévu de réaliser des analyses sur d'anciens échantillons biologiques prélevés dans le cadre de la pratique courante et conservés depuis, l'accord décrit du patient devra être obtenu pour effectuer ces analyses et pour accéder à son dossier médical (s'il n'est plus suivi pour la pathologie ayant induit les prélèvements).

10.3. Aspect réglementaire.

Les informations concernant les patients (données issues du dossier médical et enregistrements vidéo) participant à cette étude seront anonymisées (identification par le monogramme et le n° d'inclusion) et informatisées. L'avis du CPP sera demandé avant de débuter la recherche. Un résumé de la recherche sera envoyé à l'ANSM pour information. Cette étude entre bien dans la méthode de référence MR03 de la CNIL.

10.4. Rapport bénéfices/risques.

Les patients participant à cette étude n'en retireront aucun bénéfice mais n'encourront aucun risque. Les examens pratiqués ne seront pas invasifs.

XI - TRAITEMENT DES DONNÉES (JUSTIFICATION, DÉCLARATION CNIL) ET CONSERVATION DES DOCUMENTS ET DES DONNÉES RELATIVES À LA RECHERCHE

11.1. Justification du recueil des données et du type de déclaration CNIL

Les données recueillies dans le cadre de cette étude le sont dans un but de recherche scientifique, à des fins d'intérêt public.

Cette étude entre dans le cadre de la méthodologie de référence MR03 de la CNIL enregistrée, pour le CHU de St-Etienne, sous le n°1987446, pour les raisons suivantes :

- le recueil de données de santé à des fins de recherche
- l'obtention de l'avis d'un CPP avant de débuter la recherche
- l'utilisation de données anonymisées (identification par le monogramme et le n° d'inclusion)
- une information individuelle des personnes concernées
- un accès aux données uniquement par les professionnels (de santé et du promoteur) impliquées dans l'étude

Le fait que cette étude entre dans le cadre de la MR03 ainsi que les raisons seront notifiées dans le registre des traitements du promoteur.

11.2. Circuit et sécurité des données

Les données seront d'abords recueillies sur un cahier d'observation papier. Puis, la saisie des données sera réalisée, sous la responsabilité de l'investigateur, par un membre de son équipe (TEC, infirmière,…) sur un fichier Access ou Excel, hébergé sur le réseau du CHU. L'accès sera protégé par un identifiant et un mot de passe, propre à chaque personne, d'au moins 8 caractères alpha-numériques qui devra être changé tous les 3 mois. Les données sont sauvegardées de façon permanente.

Une extraction des données sous forme ASCII, TXT, Access ou SAS sera demandée pour l'analyse, afin de pouvoir faire l'importation dans le logiciel d'analyse statistique adéquate.

Il n'est prévu aucun transfert des données hors de l'Union européenne.

11.3. Modalités de traitement, vérification et validation des données (data management)

Aucun contrôle des données n'est prévu au moment de la saisie. Le data-management sera réalisé avant l'analyse statistique de façon succincte et automatique pour les contrôles de bornes relatives (données à vérifier) et absolues (données aberrantes) et les données manquantes. Des demandes de correction seront établies par variable (avec la liste des patients).

En fonction du cahier des charges, une validation des données sera éventuellement réalisée pour l'analyse statistique, et des demandes de correction seront émises à l'investigateur ou au TEC de l'étude, qui s'engage à compéter et corriger les données en conséquence.

Le gel de base sera décidé d'un commun accord entre le statisticien de l'étude, l'investigateur principal et le chef de projet.

11.4. Archivage des documents de l'essai (papier et informatique)

Les données des patients (dossiers médicaux et enregistrements vidéo) pourront être conservées jusqu'à deux ans après la dernière publication des résultats de la recherche ou, en cas d'absence de publication, jusqu'à la signature du rapport final de la recherche.

Conformément à l'arrêté du 11 août 2008 fixant la durée de conservation par le promoteur et l'investigateur des documents et données relatifs à une recherche biomédicale autre que celle portant sur des médicaments à usage humain, à la fin de la recherche :

- l'ensemble des documents (différentes versions du protocole, cahiers d'observation, classeur investigateur, consentements, correspondances,…) figurant sur support papier seront archivés, dans chaque centre, et chez le promoteur, durant au moins 15 ans.
- les données figurant sur support informatique seront conservées jusqu'à la réalisation du rapport final de la recherche puis archivées durant au moins 15 ans.

XII - PROPRIÉTÉS DES DONNÉES – PUBLICATION DES RÉSULTATS

- Les données seront la propriété du promoteur. Leur accès et leur utilisation seront sous la responsabilité de l'investigateur principal.
- Le Dr GIRAUD Antoine sera responsable de la publication.
- Le CHU de Saint-Étienne sera cité en tant que promoteur de l'étude ainsi que le ou les organismes financeurs (voir mentions sur le site internet du CHU : http://www.chu-st-etienne.fr/Recherche/Pro/Publication.asp

XIII - RÉFÉRENCES

1 Blondel, B., Lelong, N., Kermarrec, M., Goffinet, F., 2012. Trends in perinatal health in France from 1995 to 2010. Results from the French National Perinatal Surveys. Journal de Gynécologie Obstétrique et Biologie de la Reproduction 41, e1–e15. https://doi.org/10.1016/j.jgyn.2012.04.014

2 Vogel, J.P., Chawanpaiboon, S., Moller, A.-B., Watananirun, K., Bonet, M., Lumbiganon, P., 2018. The global epidemiology of preterm birth. Best Practice & Research Clinical Obstetrics & Gynaecology 52, 3–12. https://doi.org/10.1016/j.bpobgyn.2018.04.003

3 Graham, H.K., Rosenbaum, P., Paneth, N., Dan, B., Lin, J.-P., Damiano, D.L., Becher, J.G., Gaebler-Spira, D., Colver, A., Reddihough, D.S., Crompton, K.E., Lieber, R.L., 2016. Cerebral palsy. Nature Reviews Disease Primers 15082. https://doi.org/10.1038/nrdp.2015.82

4 Pierrat, V., Marchand-Martin, L., Arnaud, C., Kaminski, M., Resche-Rigon, M., Lebeaux, C., Bodeau-Livinec, F., Morgan, A.S., Goffinet, F., Marret, S., Ancel, P.-Y., Group, and the E.-2 writing, 2017. Neurodevelopmental outcome at 2 years for preterm children born at 22 to 34 weeks' gestation in France in 2011: EPIPAGE-2 cohort study. BMJ 358, j3448. https://doi.org/10.1136/bmj.j3448

5 Hintz, S.R., Vohr, B.R., Bann, C.M., Taylor, H.G., Das, A., Yolton, K., Watson, V.E., Lowe, J., DeAnda, M.E., Ball, M.B., Finer, N.N., Meurs, K.P.V., Shankaran, S., Pappas, A., Barnes, P.D., Bulas, D., Newman, J.E., Wilson-Costello, D.E., Harmon, H.M., Peralta-Carcelen, M., Adams-Chapman, I., Freeman, A., Fuller, J., Vaucher, Y.E., Colaizy, T.T., Winter, S., McGowan, C., Goldstein, R.F., Higgins, R.D., 2018. Preterm Neuroimaging and School-Age Cognitive Outcomes 142, 13.

6 Huhtala, M., Korja, R., Lehtonen, L., Haataja, L., Lapinleimu, H., Rautava, P., on behalf of the PIPARI Study Group, 2012. Parental Psychological Well-Being and Behavioral Outcome of Very Low Birth Weight Infants at 3 Years. Pediatrics 129, e937–e944. https://doi.org/10.1542/peds.2011-2411

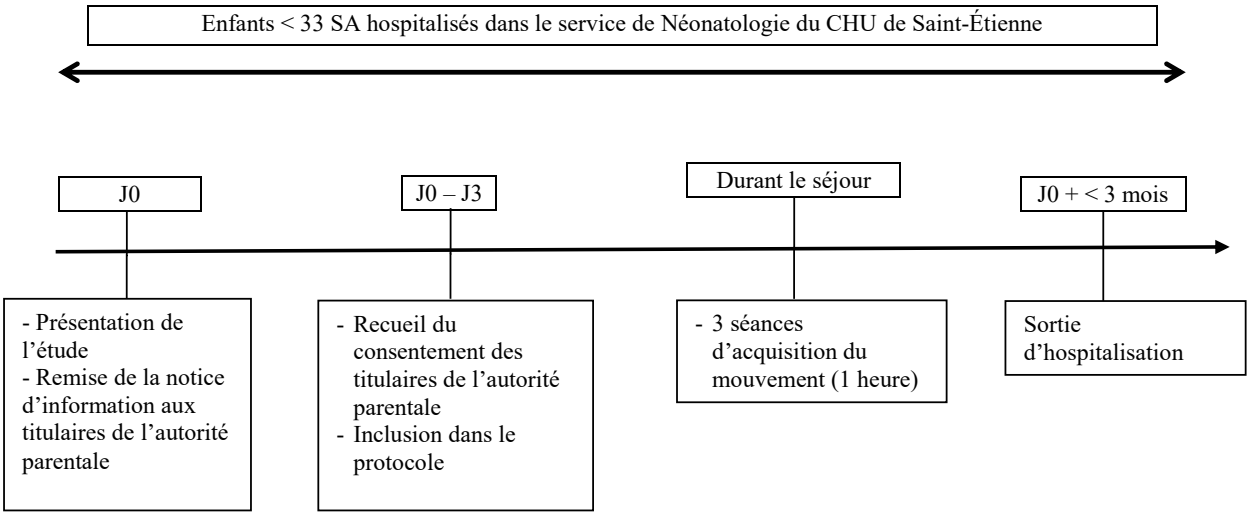
7 Einspieler, C., Prechtl, H.F.R., Bos, A., Ferrari, F., Cioni, G., 2008. Prechtl's Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants, Mac Keith Press. ed.

8 Hadders-Algra, M., 2018. Neural substrate and clinical significance of general movements: an update. Developmental Medicine & Child Neurology 60, 39–46. https://doi.org/10.1111/dmcn.13540

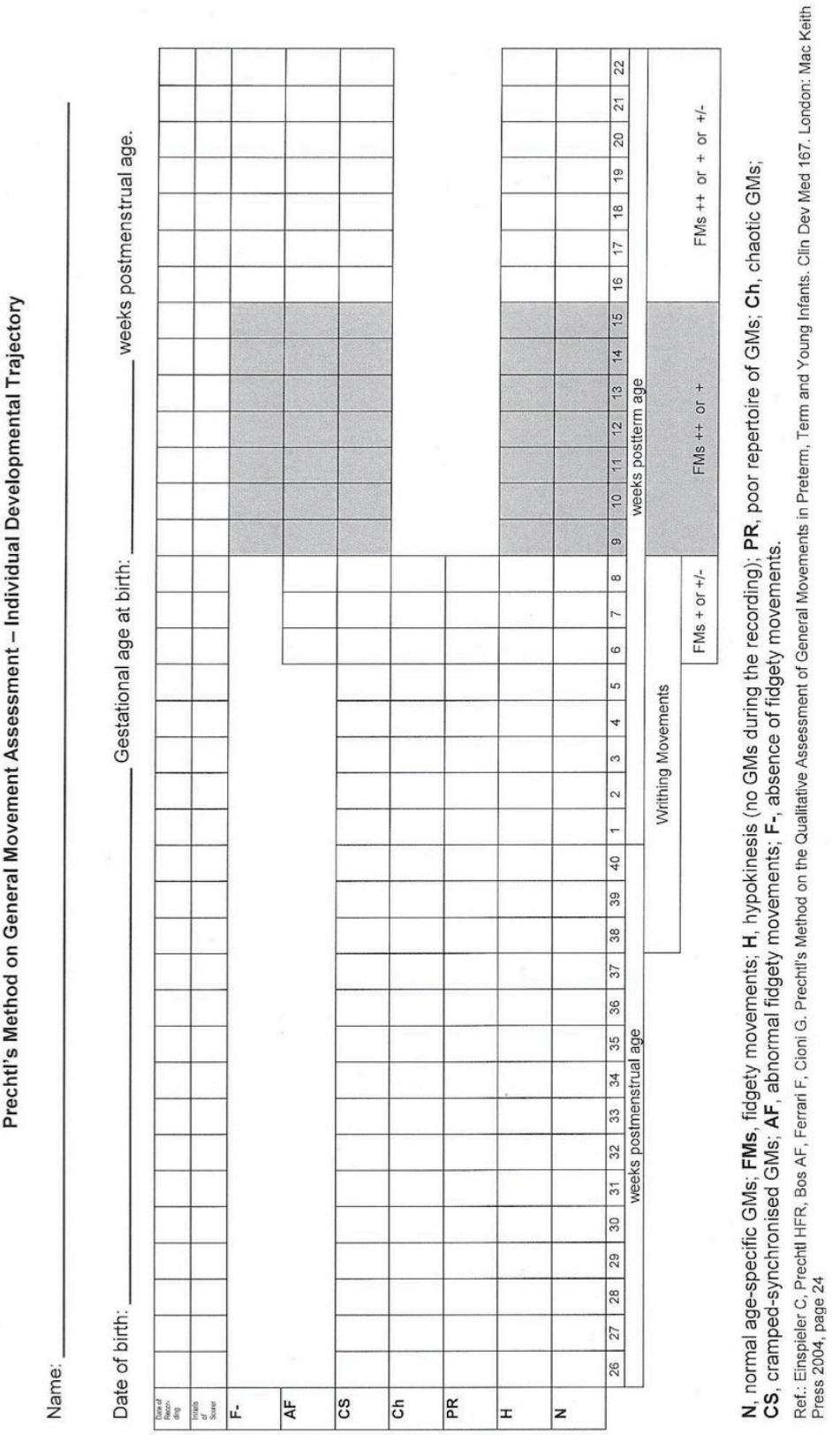
9 Giraud, A., Chaux, R., Allard, M.-J., Celle, M., Teyssier, G., Roche, F., Chapelle, C., Chabrier, S., Sébire, G., Patural, H., 2020. Perinatal inflammation is associated with social and motor impairments in preterm children without severe neonatal brain injury. European Journal of Paediatric Neurology. https://doi.org/10.1016/j.ejpn.2020.06.008

XIV - ANNEXES

ANNEXE 1 : SCHÉMA DE L'ÉTUDE



ANNEXE 2 : GRILLE DE COTATION DE LA MÉTHODE DE PRETCHL



ANNEXE 3 : AUTORISATION DE PRISE DE VUE

Sujet :.....

AUTORISATION CHU
AUTORISATION DE PRISES DE VUE ET D'INTERVIEWS

Vu l'article 9 du Code civil relatif au respect de la vie privée ;
Vu l'article L.1110-4 du Code de la santé publique relatif au respect de la vie privée et au secret des informations concernant le patient ;
Vu l'article L.1112-1 du Code de la santé publique relatif au secret médical et à la protection de la confidentialité des informations ;
Vu l'article R1112-47 du Code de la santé publique relatif à l'accès des journalistes et visiteurs aux malades ;
Vu la Circulaire DHOS du 2 mars 2006 relative aux droits des personnes hospitalisées et comportant une charte de la personne hospitalisée (notamment dans son article 9) ;

Je soussigné(e) : …… né(e) le……

☐ patient ☐ personnel ☑ représentant légal

demeurant :……

Adresse mail …….
autorise : le Centre Hospitalier Universitaire de Saint-Etienne

à : ☑ filmer ; ☐ photographier ; ☐ interviewer ;

et à utiliser l'image de :
☐ ma personne
☑ l'enfant …… ……..…...dont je suis le représentant légal
☐ le majeur protégé ……..... ……..dont je suis le représentant légal

En outre, dans le cadre de la communication générale de l'établissement, j'autorise, gracieusement et sans contrepartie, le CHU de Saint-Étienne à utiliser et diffuser les prises de vues photographiques / filmées qui ont été prises, dans* :

Le journal interne
Les sites internet et intranet
Les réseaux sociaux (Facebook, Twitter…etc.)
Les documents institutionnels remis aux patients et usagers

Les documents pédagogiques diffusés lors de congrès, cours, colloques, conférence (

La presse …….. dans le cadre exclusif du sujet exposé ci-avant
Autre…….

Fait à Saint- Etienne le……,

Signature précédée de la mention « lu et approuvé »

Merci **retourner** ce document dûment complété et **signé à**
CHU de Saint-Etienne, Direction Générale –Service Communication Hôpital Bellevue

Attestation

Service de REANIMATION PEDIATRIQUE ET NEONATOLOGIE
HÔPITAL NORD
Avenue Albert Raimond
Saint Priest en Jarez
Téléphone 04 77 82 81 17
Télécopie 04 77 82 84 54

Professeur H. PATURAL
E -mail : hugues.patural@chu-st-etienne.fr

EA SNA Epis- 4607 – Université de Lyon Saint-Etienne

Saint Etienne,
Le 6 septembre 2020

421039991

Je soussigné Professeur Hugues PATURAL, atteste que Monsieur Ameur SOUALMI né le 14 Avril 1996, bénéficie d'un contrat doctoral à partir du 1er Octobre 2020 au sein de l'école doctorale EDSIS Science Ingénierie Santé de l'université de Saint-Etienne. Il bénéficiera à ce titre d'une enveloppe budgétaire de 60000 Euros de l'EDSIS, répartis sur 3 années à partir du 01/10/2020 et travaillera au sein du laboratoire SNA 4607 de l'UJM à compter de cette date.

Attestation remise à l'intéressé pour faire valoir.

Professeur H. PATURAL

A 314 C

Accord du Directeur de Laboratoire

Service de PHYSIOLOGIE CLINIQUE ET DE L'EXERCICE
Centre VISAS
VIeillissement et **S**yndrome d'**A**pnée du **S**ommeil
Promotion de l'Activité Physique
Hôpital NORD
Bâtiment A-3è étage
Avenue Albert Raimond
St- Priest en Jarez

Professeur F. ROCHE
Cardiologue
Chef de service
RPPS 10003013561

Professeur JC BARTHELEMY
Cardiologue
RPPS 10003008066

Docteur D.HUPIN
Médecin du Sport
RPPS 10100531937

Docteur PB.POBLE
Pneumologue

Docteur S.CHOMETTE BALLEREAU
Exploration fonctionnelle respiratoire et pathologie du sommeil
RPPS 10100198141

Secrétariat VISAS
8h30-12h30
Tel : 04 77 82 91 09
Fax : 04 77 82 91 10
visas.secretariat@chu-st-etienne.fr

Saint-Etienne, le 14/09/2020

Je soussigné Pr Frédéric Roche atteste soutenir le projet de recherche intitulé « Développement d'un outil pronostique développemental de l'enfant prématuré basé sur l'analyse automatisée de la motricité spontanée » porté par le Dr Antoine Giraud.

En tant que directeur de laboratoire, j'atteste être en mesure d'assurer l'ensemble des frais de fonctionnement en rapport avec ce projet de recherche, aujourd'hui estimés à 5 000 euros.

Bien respectueusement,

Pr Frédéric Roche
Physiologie Clinique, CHU Nord
Directeur de l'EA 4607 SNA EPIS
UJM Saint Etienne, UdL

N° FINESS
420785878

B

Bibliography

Outline

References

[1] A. Giraud, “Impact de l’inflammation périnatale sur le cerveau en développement : de la dimension neurodéveloppementale à l’échelle endothéliale,” Theses, Université de Lyon, Nov. 2020.

[2] G. Cioni, F. Ferrari, A. F. Bos, H. F. R. Prechtl, and C. Einspieler, *Prechtls Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants*. Mac Keith Press, 2008.

[3] S. Reich, D. Zhang, T. Kulvicius, *et al.*, “Novel ai driven approach to classify infant motor functions,” *Scientific Reports*, vol. 11, May 2021. DOI: 10.1038/s41598-021-89347-5.

[4] K. D. McCay, P. Hu, H. P. H. Shum, *et al.*, “A pose-based feature fusion and classification framework for the early prediction of cerebral palsy in infants,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 8–19, 2022. DOI: 10.1109/TNSRE.2021.3138185.

[5] D. Sakkos, K. McCay, C. Marcroft, N. Embleton, S. Chattopadhyay, and E. Ho, “Identification of abnormal movements in infants: a deep neural network for body part-based prediction of cerebral palsy,” *IEEE Access*, vol. PP, pp. 1–1, Jun. 2021. DOI: 10.1109/ACCESS.2021.3093469.

[6] I. Doroniewicz, D. Ledwoń, A. Affanasowicz, *et al.*, “Writhing movement detection in newborns on the second and third day of life using pose-based feature machine learning classification,” *Sensors*, vol. 20(21), 5986, Oct. 2020. DOI: 10.3390/s20215986.

[7] C. Chambers, N. Seethapathi, R. Saluja, *et al.*, “Computer vision to automatically assess infant neuromotor risk,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, pp. 2431–2442, Nov. 2020. DOI: 10.1109/TNSRE.2020.3029121.

[8] N. Hesse, C. Bodensteiner, M. Arens, U. G. Hofmann, R. 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*, L. Leal-Taixé and S. Roth, Eds., Cham: Springer International Publishing, 2019, pp. 32–49, ISBN: 978-3-030-11024-6.

[9] X. Huang, N. Fu, S. Liu, and S. Ostadabbas, "Invariant representation learning for infant pose estimation with small data," in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, Los Alamitos, CA, USA: IEEE Computer Society, Dec. 2021, pp. 1–8. DOI: 10.1109/FG52635.2021.9666956.

[10] J.-M. Moutquin, "Classification and heterogeneity of preterm birth," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 110, 2003.

[11] "The european perinatal health report, 2015-2019."

[12] R. M. Patel, S. C. Kandefer, M. C. Walsh, *et al.*, "Causes and timing of death in extremely premature infants from 2000 through 2011.," *Obstetric Anesthesia Digest*, 2015.

[13] V. Pierrat, L. Marchand-Martin, S. Marret, *et al.*, "Neurodevelopmental outcomes at age 5 among children born preterm: epipage-2 cohort study," *BMJ*, vol. 373, 2021. DOI: 10.1136/bmj.n741. eprint: https://www.bmj.com/content/373/bmj.n741.full.pdf.

[14] G. S. Berkowitz and E. Papiernik, "Epidemiology of Preterm Birth," *Epidemiologic Reviews*, vol. 15, no. 2, pp. 414–443, Jul. 1993, ISSN: 0193-936X. DOI: 10.1093/oxfordjournals.epirev.a036128.

[15] L. M. Savitz David A. & Pastore, *Causes of Prematurity*. Cambridge, England: Cambridge University Press, 1999, pp. 63–104.

[16] M. Nordentoft, H. C. Lou, D. Hansen, *et al.*, "Intrauterine growth retardation and premature delivery: the influence of maternal smoking and psychosocial factors.," *American Journal of Public Health*, vol. 86, no. 3, pp. 347–354, 1996. DOI: 10.2105/AJPH.86.3.347.

[17] U. Kesmodel, S. Olsen, and N. Secher, "Does alcohol increase the risk of preterm delivery?" Dansk, *Epidemiology*, vol. 11, no. 5, pp. 512–518, 2000, ISSN: 1044-3983.

[18] R. Behrman and A. Butler, *Preterm birth: Causes, Consequences, and prevention*. May 2007, pp. 1–772. DOI: 10.17226/11622.

[19] M. Hedegaard, T. B. Henriksen, N. J. Secher, M. C. Hatch, and S. Sabroe, "Do stressful life events affect duration of gestation and risk of preterm delivery?" *Epidemiology*, vol. 7, pp. 339–345, 1996.

[20] E. L. Davies, J. S. Bell, and S. Bhattacharya, “Preeclampsia and preterm delivery: a population-based case–control study,” *Hypertension in Pregnancy*, vol. 35, no. 4, pp. 510–519, 2016. DOI: 10.1080/10641955.2016.1190846.

[21] G. R. Alexander, M. D. Kogan, D. Bader, W. Carlo, M. C. Allen, and J. M. Mor, “Us birth weight/gestational age-specific neonatal mortality: 1995-1997 rates for whites, hispanics, and blacks.,” *Pediatrics*, vol. 111 1, e61–6, 2003.

[22] F. Lammertink, C. H. Vinkers, M. L. Tataranno, and M. J. Benders, “Premature birth and developmental programming: mechanisms of resilience and vulnerability,” *Frontiers in Psychiatry*, vol. 11, 2021.

[23] P. Rosenbaum, N. Paneth, A. Leviton, *et al.*, “A report: the definition and classification of cerebral palsy april 2006,” *Developmental medicine and child neurology. Supplement*, vol. 109, pp. 8–14, Mar. 2007. DOI: 10.1111/j.1469-8749.2007.tb12610.x.

[24] B. R. Vohr, L. L. Wright, W. K. Poole, S. A. McDonald, and for the NICHD Neonatal Research Network Follow-up Study, “Neurodevelopmental Outcomes of Extremely Low Birth Weight Infants <32 Weeks’ Gestation Between 1993 and 1998,” *Pediatrics*, vol. 116, no. 3, pp. 635–643, Sep. 2005, ISSN: 0031-4005. DOI: 10.1542/peds.2004-2247. eprint: https://publications.aap.org/pediatrics/article-pdf/116/3/635/1005066/zpe00905000635.pdf.

[25] L. Doyle, “Outcome at 5 years of age of children 23 to 27 weeks’ gestation: refining the prognosis.,” *Pediatrics*, vol. 108 1, pp. 134–41, 2001.

[26] D. Elbourne, S. Ayers, H. D. Dellagrammaticas, A. H. Johnson, M. Leloup, and S. Lenoir-Piat, “Randomised controlled trial of prophylactic etamsylate: follow up at 2 years of age,” *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 84, F183–F187, 2001.

[27] S. L. Carter, *Motor impairment associated with neurological injury in premature infants*.

[28] R. E. Grunau, M. F. Whitfield, and C. Davis, “Pattern of learning disabilities in children with extremely low birth weight and broadly average intelligence.,” *Archives of pediatrics & adolescent medicine*, vol. 156 6, pp. 615–20, 2002.

[29] D. Scholes, T. M. Hooton, P. L. Roberts, K. Gupta, A. E. Stapleton, and W. E. Stamm, “Risk factors associated with acute pyelonephritis in healthy women,” *Annals of Internal Medicine*, vol. 142, pp. 20–27, 2005.

[30] I. L. Hand, R. A. Shellhaas, S. S. Milla, *et al.*, “Routine Neuroimaging of the Preterm Brain,” *Pediatrics*, vol. 146, no. 5, e2020029082, Nov. 2020, ISSN: 0031-4005. DOI: 10.1542/peds.2020-029082. eprint: https://publications.

aap.org/pediatrics/article-pdf/146/5/e2020029082/1081278/peds_2020029082.pdf.

[31] G. Variane, D. Rodrigues, R. Pietrobom, C. França, A. Netto, and M. Magalhaes, "Newborns at high risk for brain injury: the role of the amplitude-integrated electroencephalography," *Jornal de Pediatria*, vol. 98, Jan. 2022. DOI: 10.1016/j.jped.2021.10.008.

[32] M. Bosanquet, L. Copeland, R. Ware, and R. Boyd, "A systematic review of tests to predict cerebral palsy in young children," *Developmental medicine and child neurology*, vol. 55, pp. 418–26, May 2013. DOI: 10.1111/dmcn.12140.

[33] A. J. Spittle, L. W. Doyle, and R. N. Boyd, "A systematic review of the clinimetric properties of neuromotor assessments for preterm infants during the first year of life," *Developmental Medicine & Child Neurology*, vol. 50, 2008.

[34] H. F. Prechtl and B. Hopkins, "Developmental transformations of spontaneous movements in early infancy," *Early Human Development*, vol. 14, no. 3, pp. 233–238, 1986, ISSN: 0378-3782. DOI: https://doi.org/10.1016/0378-3782(86)90184-2.

[35] C. Einspieler, A. F. Bos, M. E. Libertus, and P. B. Marschik, "The general movement assessment helps us to identify preterm infants at risk for cognitive dysfunction," *Frontiers in Psychology*, vol. 7, 2016, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2016.00406.

[36] M. Hadders-Algra, A. G. Boxum, T. Hielkema, and E. G. Hamer, "Effect of early intervention in infants at very high risk of cerebral palsy: a systematic review," *Developmental Medicine & Child Neurology*, vol. 59, no. 3, pp. 246–258, DOI: https://doi.org/10.1111/dmcn.13331. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/dmcn.13331.

[37] H. Prechtl, "Qualitative changes of spontaneous movements in fetus and preterm infant are a marker of neurological dysfunction," *Early Human Development*, vol. 23, no. 3, pp. 151–158, 1990, New Studies on Movement Assessment in Fetuses and Preterm Infants, ISSN: 0378-3782. DOI: https://doi.org/10.1016/0378-3782(90)90011-7.

[38] G. Cioni and H. F. Prechtl, "Preterm and early postterm motor behaviour in low-risk premature infants," *Early Human Development*, vol. 23, no. 3, pp. 159–191, 1990, New Studies on Movement Assessment in Fetuses and Preterm Infants, ISSN: 0378-3782. DOI: https://doi.org/10.1016/0378-3782(90)90012-8.

[39] H. F. Prechtl, C. Einspieler, G. Cioni, A. F. Bos, F. Ferrari, and D. Sontheimer, "An early marker for neurological deficits after perinatal brain lesions," *The*

Lancet, vol. 349, no. 9062, pp. 1361–1363, 1997, ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(96)10182-3.

[40] F. Ferrari, G. Cioni, and H. Prechtl, "Qualitative changes of general movements in preterm infants with brain lesions," *Early Human Development*, vol. 23, no. 3, pp. 193–231, 1990, New Studies on Movement Assessment in Fetuses and Preterm Infants, ISSN: 0378-3782. DOI: https://doi.org/10.1016/0378-3782(90)90013-9.

[41] C. Einspieler and H. F. R. Prechtl, "Prechtl's assessment of general movements: a diagnostic tool for the functional assessment of the young nervous system," *Mental Retardation and Developmental Disabilities Research Reviews*, vol. 11, no. 1, pp. 61–67, DOI: https://doi.org/10.1002/mrdd.20051. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrdd.20051.

[42] C. Einspieler, H. F. Prechtl, F. Ferrari, G. Cioni, and A. F. Bos, "The qualitative assessment of general movements in preterm, term and young infants — review of the methodology," *Early Human Development*, vol. 50, no. 1, pp. 47–60, 1997, Spontaneous Motor Activity as a Diagnostic Tool Functional Assessment of the Young Nervous System, ISSN: 0378-3782. DOI: https://doi.org/10.1016/S0378-3782(97)00092-3.

[43] T. Valentin, K. Uhl, and C. Einspieler, "The effectiveness of training in prechtl's method on the qualitative assessment of general movements," *Early Human Development*, vol. 81, no. 7, pp. 623–627, 2005, ISSN: 0378-3782. DOI: https://doi.org/10.1016/j.earlhumdev.2005.04.003.

[44] Y. Noble and R. Boyd, "Neonatal assessments for the preterm infant up to 4months corrected age: a systematic review," *Developmental medicine and child neurology*, vol. 54, pp. 129–39, Dec. 2011. DOI: 10.1111/j.1469-8749.2010.03903.x.

[45] S. Aghanavesi, F. Bergquist, D. Nyholm, M. Senek, and M. Memedi, "Motion sensor-based assessment of parkinson's disease motor symptoms during leg agility tests: results from levodopa challenge," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 111–119, 2020. DOI: 10.1109/JBHI.2019.2898332.

[46] A. M. Amiri, N. Peltier, C. Goldberg, *et al.*, "Wearsense: detecting autism stereotypic behaviors through smartwatches," *Healthcare*, vol. 5, no. 1, 2017, ISSN: 2227-9032. DOI: 10.3390/healthcare5010011.

[47] D. Gravem, M. Singh, C. Chen, *et al.*, "Assessment of Infant Movement With a Compact Wireless Accelerometer System," *Journal of Medical Devices*, vol. 6, no. 2, May 2012, 021013, ISSN: 1932-6181. DOI: 10.1115/1.4006129.

[48] M. Singh and D. J. Patterson, "Involuntary gesture recognition for predicting cerebral palsy in high-risk infants," in *International Symposium on Wearable Computers (ISWC) 2010*, 2010, pp. 1–8. DOI: 10.1109/ISWC.2010.5665873.

[49] F. Heinze, N. Breitbach-Faller, T. Schmitz-Rode, and C. Disselhorst-Klug, "Movement analysis by accelerometry of newborns for the early detection of movement disorders due to infantile cerebral palsy," in *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, O. Dössel and W. C. Schlegel, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 24–27, ISBN: 978-3-642-03889-1.

[50] D. Karch, K.-S. Kim, K. Wochner, J. Pietz, H. Dickhaus, and H. Philippi, "Quantification of the segmental kinematics of spontaneous infant movements.," *Journal of biomechanics*, vol. 41 13, pp. 2860–7, 2008.

[51] H. Philippi, D. Karch, K.-S. Kang, *et al.*, "Computer-based analysis of general movements reveals stereotypies predicting cerebral palsy," *Developmental Medicine & Child Neurology*, vol. 56, no. 10, pp. 960–967, 2014. DOI: https://doi.org/10.1111/dmcn.12477. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/dmcn.12477.

[52] 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*, vol. 25, pp. 125–44, May 2006. DOI: 10.1016/j.humov.2005.09.012.

[53] N. Kanemaru, H. Watanabe, H. Kihara, *et al.*, "Jerky spontaneous movements at term age in preterm infants who later developed cerebral palsy," *Early Human Development*, vol. 90, no. 8, pp. 387–392, 2014, ISSN: 0378-3782. DOI: https://doi.org/10.1016/j.earlhumdev.2014.05.004.

[54] 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*, vol. 25, no. 2, pp. 125–144, 2006, ISSN: 0167-9457. DOI: https://doi.org/10.1016/j.humov.2005.09.012.

[55] L. Berthouze and M. Mayston, "Design and validation of surface-marker clusters for the quantification of joint rotations in general movements in early infancy,"

Journal of Biomechanics, vol. 44, no. 6, pp. 1212–1215, 2011, ISSN: 0021-9290. DOI: https://doi.org/10.1016/j.jbiomech.2011.01.016.

[56] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, and R. Støen, "Using computer-based video analysis in the study of fidgety movements," *Early Human Development*, vol. 85, no. 9, pp. 541–547, 2009, ISSN: 0378-3782. DOI: https://doi.org/10.1016/j.earlhumdev.2009.05.003.

[57] A. Stahl, C. Schellewald, Ø. Stavdahl, O. Aamo, L. Adde, and H. Kirkerød, "An optical flow-based method to predict infantile cerebral palsy," *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 20, pp. 605–14, Apr. 2012. DOI: 10.1109/TNSRE.2012.2195030.

[58] S. Orlandi, K. Raghuram, C. R. Smith, *et al.*, "Detection of atypical and typical infant movements using computer-based video analysis," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 3598–3601. DOI: 10.1109/EMBC.2018.8513078.

[59] K. Raghuram, S. Orlandi, V. Shah, *et al.*, "Automated movement analysis to predict motor impairment in preterm infants: a retrospective study," *Journal of Perinatology*, vol. 39, pp. 1–8, Oct. 2019. DOI: 10.1038/s41372-019-0464-0.

[60] H. Rahmati, O. M. Aamo, Ø. Stavdahl, R. Dragon, and L. Adde, "Video-based early cerebral palsy prediction using motion segmentation," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 3779–3783. DOI: 10.1109/EMBC.2014.6944446.

[61] W. Schmidt, M. Regan, M. Fahey, and A. Paplinski, "General movement assessment by machine learning: why is it so difficult?" *Journal of Medical Artificial Intelligence*, vol. 2, no. 0, 2019, ISSN: 2617-2496.

[62] T. Tsuji, S. Nakashima, H. Hayashi, *et al.*, "Markerless measurement and evaluation of general movements in infants," *Scientific Reports*, vol. 10, Jan. 2020. DOI: 10.1038/s41598-020-57580-z.

[63] T.-Y. Lin, M. Maire, S. Belongie, *et al.*, "Microsoft coco: common objects in context," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham: Springer International Publishing, 2014, pp. 740–755, ISBN: 978-3-319-10602-1.

[64] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: new benchmark and state of the art analysis," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3686–3693. DOI: 10.1109/CVPR.2014.471.

[65] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017.

[66] Z. Geng, K. Sun, B. Xiao, Z. Zhang, and J. Wang, “Bottom-up human pose estimation via disentangled keypoint regression,” *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14 671–14 681, 2021.

[67] S. Jin, W. Liu, E. Xie, *et al.*, “Differentiable hierarchical graph grouping for multi-person pose estimation,” in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., Cham: Springer International Publishing, 2020, pp. 718–734.

[68] S. Kreiss, L. Bertoni, and A. Alahi, “Pifpaf: composite fields for human pose estimation,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11 969–11 978.

[69] Z. Luo, Z. Wang, Y. Huang, L. Wang, T. Tan, and E. Zhou, “Rethinking the heatmap regression for bottom-up human pose estimation,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 13 259–13 268.

[70] G. Papandreou, T. Zhu, L.-C. Chen, S. Gidaris, J. Tompson, and K. Murphy, “Personlab: person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model,” in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Cham: Springer International Publishing, 2018, pp. 282–299.

[71] A. Nibali, Z. He, S. Morgan, and L. Prendergast, “3d human pose estimation with 2d marginal heatmaps,” in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Los Alamitos, CA, USA: IEEE Computer Society, Jan. 2019, pp. 1477–1485. DOI: 10.1109/WACV.2019.00162.

[72] A. Newell, Z. Huang, and J. Deng, “Associative embedding: end-to-end learning for joint detection and grouping,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS’17, Long Beach, California, USA: Curran Associates Inc., 2017, pp. 2274–2284, ISBN: 9781510860964.

[73] G. Rogez, P. Weinzaepfel, and C. Schmid, “Lcr-net: localization-classification-regression for human pose,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1216–1224. DOI: 10.1109/CVPR.2017.134.

[74] U. Iqbal and J. Gall, “Multi-person pose estimation with local joint-to-person associations,” in *Computer Vision – ECCV 2016 Workshops*, G. Hua and H. Jégou, Eds., Cham: Springer International Publishing, 2016, pp. 627–642, ISBN: 978-3-319-48881-3.

[75] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang, and L. Zhang, “Higherhrnet: scale-aware representation learning for bottom-up human pose estimation,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5385–5394. DOI: 10.1109/CVPR42600.2020.00543.

[76] C. Zheng, W. Wu, T. Yang, *et al.*, *Deep learning-based human pose estimation: a survey*, Dec. 2020.

[77] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, “Cascaded pyramid network for multi-person pose estimation,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7103–7112.

[78] B. Xiao, H. Wu, and Y. Wei, “Simple baselines for human pose estimation and tracking,” in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Cham: Springer International Publishing, 2018, pp. 472–487.

[79] G. Lin, A. Milan, C. Shen, and I. Reid, “Refinenet: multi-path refinement networks for high-resolution semantic segmentation,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5168–5177.

[80] A. Newell, K. Yang, and J. Deng, “Stacked hourglass networks for human pose estimation,” in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 483–499.

[81] K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep high-resolution representation learning for human pose estimation,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5686–5696. DOI: 10.1109/CVPR.2019.00584.

[82] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille, and X. Wang, “Multi-context attention for human pose estimation,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, Jul. 2017, pp. 5669–5678. DOI: 10.1109/CVPR.2017.601.

[83] F. Zhang, X. Zhu, H. Dai, M. Ye, and C. Zhu, “Distribution-aware coordinate representation for human pose estimation,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7091–7100. DOI: 10.1109/CVPR42600.2020.00712.

[84] G. Sciortino, G. M. Farinella, S. Battiato, M. Leo, and C. Distante, "On the estimation of children's poses," in *Image Analysis and Processing - ICIAP 2017*, S. Battiato, G. Gallo, R. Schettini, and F. Stanco, Eds., Cham: Springer International Publishing, 2017, pp. 410–421, ISBN: 978-3-319-68547-2. DOI: 10.1007/978-3-319-68548-9_38.

[85] S. Moccia, L. Migliorelli, R. Pietrini, and E. Frontoni, "Preterm infants' limb-pose estimation from depth images using convolutional neural networks," in *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2019, pp. 1–7. DOI: 10.1109/CIBCB.2019.8791242.

[86] D. Groos, H. Ramampiaro, and E. Ihlen, "Efficientpose: scalable single-person pose estimation," *Applied Intelligence*, vol. 51, pp. 2518–2533, 2021. DOI: 10.1007/s10489-020-01918-7.

[87] 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 *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 5469–5472. DOI: 10.1109/EMBC.2019.8857680.

[88] 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*, vol. 8, pp. 51 582–51 592, 2020. DOI: 10.1109/ACCESS.2020.2980269.

[89] S. Li, W. Zhang, and A. B. Chan, "Maximum-margin structured learning with deep networks for 3d human pose estimation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2848–2856.

[90] B. Tekin, I. Katircioglu, M. Salzmann, V. Lepetit, and P. Fua, "Structured prediction of 3d human pose with deep neural networks," Sep. 2016. DOI: 10.5244/C.30.130.

[91] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Coarse-to-fine volumetric prediction for single-image 3d human pose," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1263–1272.

[92] D. C. Luvizon, H. Tabia, and D. Picard, "Human pose regression by combining indirect part detection and contextual information," *Computers Graphics*, vol. 85, pp. 15–22, 2019, ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2019.09.002.

[93] X. Sun, B. Xiao, S. Liang, and Y. Wei, "Integral human pose regression," *ArXiv*, vol. abs/1711.08229, 2017.

[94] J. Martinez, R. Hossain, J. Romero, and J. J. Little, “A simple yet effective baseline for 3d human pose estimation,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2659–2668. DOI: 10.1109/ICCV.2017.288.

[95] K. Zhou, X. Han, N. Jiang, K. Jia, and J. Lu, “Hemlets pose: learning part-centric heatmap triplets for accurate 3d human pose estimation,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 2344–2353.

[96] F. Moreno-Noguer, “3d human pose estimation from a single image via distance matrix regression,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1561–1570.

[97] C. Li and G. H. Lee, “Generating multiple hypotheses for 3d human pose estimation with mixture density network,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9879–9887. DOI: 10.1109/CVPR.2019.01012.

[98] B. Tekin, P. Marquez-Neila, M. Salzmann, and P. Fua, “Learning to fuse 2d and 3d image cues for monocular body pose estimation,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3961–3970.

[99] S. Amin, M. Andriluka, M. Rohrbach, and B. Schiele, “Multi-view pictorial structures for 3d human pose estimation,” in *British Machine Vision Conference*, 2013.

[100] M. Bergtholdt, J. Kappes, S. Schmidt, and C. Schnörr, “A study of parts-based object class detection using complete graphs,” *International Journal of Computer Vision*, vol. 87, pp. 93–117, Mar. 2010. DOI: 10.1007/s11263-009-0209-1.

[101] M. Burenius, J. Sullivan, and S. Carlsson, “3d pictorial structures for multiple view articulated pose estimation,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3618–3625. DOI: 10.1109/CVPR.2013.464.

[102] V. Belagiannis, S. Amin, M. Andriluka, B. Schiele, N. Navab, and S. Ilic, “3d pictorial structures for multiple human pose estimation,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1669–1676. DOI: 10.1109/CVPR.2014.216.

[103] M. Kocabas, S. Karagoz, and E. Akbas, “Self-supervised learning of 3d human pose using multi-view geometry,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1077–1086, 2019.

[104] K. Iskakov, E. Burkov, V. Lempitsky, and Y. Malkov, “Learnable triangulation of human pose,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 7717–7726. DOI: 10.1109/ICCV.2019.00781.

[105] H. Qiu, C. Wang, J. Wang, N. Wang, and W. Zeng, “Cross view fusion for 3d human pose estimation,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Los Alamitos, CA, USA: IEEE Computer Society, Nov. 2019, pp. 4341–4350. DOI: 10.1109/ICCV.2019.00444.

[106] Q. Wu, G. Xu, F. Wei, C. Longting, and S. Zhang, “Rgb-d videos-based early prediction of infant cerebral palsy via general movements complexity,” *IEEE Access*, vol. PP, pp. 1–1, Mar. 2021. DOI: 10.1109/ACCESS.2021.3066148.

[107] M. Li, F. Wei, Y. Li, S. Zhang, and G. Xu, “Three-dimensional pose estimation of infants lying supine using data from a kinect sensor with low training cost,” *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6904–6913, 2021. DOI: 10.1109/JSEN.2020.3037121.

[108] Q. Wu, G. Xu, S. Zhang, Y. Li, and F. Wei, “Human 3d pose estimation in a lying position by rgb-d images for medical diagnosis and rehabilitation,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 5802–5805. DOI: 10.1109/EMBC44109.2020.9176407.

[109] S. S. Shivakumar, H. Loeb, D. K. Bogen, *et al.*, “Stereo 3d tracking of infants in natural play conditions,” in *2017 International Conference on Rehabilitation Robotics (ICORR)*, 2017, pp. 841–846. DOI: 10.1109/ICORR.2017.8009353.

[110] W. Baccinelli, M. Bulgheroni, V. Simonetti, *et al.*, “Movidea: a software package for automatic video analysis of movements in infants at risk for neurodevelopmental disorders,” *Brain Sciences*, vol. 10, 2020.

[111] H. Bridge, I. B. Ip, and A. J. Parker, “Investigating the human binocular visual system using multi-modal magnetic resonance imaging,” *Perception*, vol. 52, no. 7, pp. 441–458, 2023, PMID: 37272064. DOI: 10.1177/03010066231178664. eprint: https://doi.org/10.1177/03010066231178664.

[112] J. West, D. Ventura, and S. Warnick, “Spring research presentation: a theoretical foundation for inductive transfer,” *Brigham Young University, College of Physical and Mathematical Sciences*, vol. 1, no. 08, 2007.

[113] L. Alzubaidi, J. Zhang, A. J. Humaidi, *et al.*, “Review of deep learning: concepts, cnn architectures, challenges, applications, future directions,” *Journal of Big Data*, vol. 8, 2021.

[114] J. Wang, K. Sun, T. Cheng, *et al.*, “Deep high-resolution representation learning for visual recognition,” *IEEE Transactions on Pattern Analysis amp; Machine Intelligence*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021, ISSN: 1939-3539. DOI: 10.1109/TPAMI.2020.2983686.

[115] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, “Joint training of a convolutional network and a graphical model for human pose estimation,” in *NIPS*, 2014.

[116] *Coco keypoint evaluation*, https://cocodataset.org.

[117] A. Pavel, J. Rennie, L. Vries, *et al.*, “A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial,” *The Lancet Child & Adolescent Health*, vol. 4, Aug. 2020. DOI: 10.1016/S2352-4642(20)30239-X.

[118] S. Cabon, R. Weber, L. Cailleau, G. Carrault, P. Pladys, and F. Porée, “Automated quiet sleep detection for premature newborns based on video and ecg analysis,” in *2021 Computing in Cardiology (CinC)*, vol. 48, 2021, pp. 1–4. DOI: 10.23919/CinC53138.2021.9662821.

[119] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, OpenReview.net, 2017.

List of Publications

Journal Article

A. Soualmi, C. Ducottet, H. Patural, A. Giraud, and O. Alata, “A 3D pose estimation framework for preterm infants hospitalized in the Neonatal Unit”, *Multimed Tools Appl (2023),* https://doi.org/10.1007/s11042-023-16333-6.

International Conference

A. Soualmi, O. Alata, C. Ducottet, H. Patural, and A. Giraud, “Mean 3d dispersion for automatic general movement assessment of preterm infants”, *45th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2023. https://arinex.com.au/EMBC/pdf/full-paper_617.pdf

National Conference

A. Soualmi, “Automation de l’évaluation des mouvements généraux”, Oral Presentation, *Journées Francophones de Recherche en Néonatologie*, 2022.

National Conference

A. Soualmi, “A 3D pose estimation framework for infant general movement assessment”, Oral Presentation, *Journée Action « Visage, geste, action et comportement », GDR ISIS*, 2022.