DeepInMotion: Explainable artificial intelligent system to discover new infant movement biomarkers for early detection of disease

Relevance to the call

The DeepInMotion project will generate new knowledge and techniques in the research area of explainable artificial intelligence (XAI) for early detection of motor disabilities in children for clinical decision support. The project addresses important challenges for quantitative and accurate diagnosis of motor disabilities, providing a basis for planning health care services, improved prevention, and treatment strategies, as well as providing a research tool for clinical movement analysis. The new XAI techniques will be integrated in a low-cost clinical service implementation providing an easy-to-use, low-threshold, and highly available decision support system. The DeepInMotion system will ensure healthy lives and promote well-being for all children reducing inequality of health care within and among countries (UN sustainable development goal 3 and 10, see Figure 1 below). The project will utilize the world largest international database of videos of high-risk infants, administrated by St Olavs Hospital in Norway, to develop the new XAI techniques. The explanation accuracy, transparency and accountability of the AI-technology will be co-developed by an international unique interdisciplinary group of computer scientists, human movement scientists, and specialists within paediatric research and physiotherapy including two Hospital clinics in the Central Norway Regional Health Authority (St Olavs Hospital and Ålesund Hospital). The project advisory board will contain several international medical device manufacturers that will exploit the IP delivered by the present project to meet the UN sustainable goals.

1. Excellence

1.1 State of the art, knowledge needs and project objectives

Cerebral palsy (CP) is a movement disorder caused by a perinatal brain injury that results in life-long needs for special services and care. Treatment and care for individuals with CP results in a life-time cost of €800 000 per child, approximately four times that of a typically developing child. The condition also results in severe personal challenges for the child and their families [1, 2]. Today, early detection of CP is performed by a subjective and a qualitative movement analysis, called the General Movement Assessment (GMA), in infants between 12 to 18 weeks post term age [3]. GMA is currently the most accurate method for detection of CP before 5 months of age and provide opportunities for early onset of therapies and treatments in the period when plasticity of the brain is at its highest [4]. However, GMA need highly qualified clinicians with long experience to be reliable and, thus, lack of widespread adoption among clinical teams [4, 5]. GMA is also a qualitative method used in an *ad hoc* and subjective manner without objective identification of diagnostic-specific movement biomarkers. Even though quantitative methods have been suggested for early detection of CP, they are all based on pre-selected movement features for which the association to the CP outcome are unknown and, consequently, lack sufficient accuracy and robustness to be implemented in clinical services [6].

During the last decade, a large research field within computer science have emerge developing innovative machine learning models called deep neural networks (DNN). DNNs can automatically aggregate new features from big data sources resulting in super-human performances in complex decision tasks [7]. Thus, DNNs have the potential to discovering new movement features that are directly related to the CP outcome. Even though DNNs can make complex decisions with high performance, the networks are not self-explainable and does not provide the logic behind it's decisions or automatic aggregated features to the end-user [8]. In movement analysis for clinical decision support, the logic behind the automatic aggregation of features within DNNs need to be revealed through clinical meaningful explanations for the features to become biomarkers for a diagnosis or progression of disease [8]. Even though conventional machine learning models have been suggested for early detection of CP [9,10], they are all developed as "black box" models without necessary explanation of their decisions to adhere to EU-MDR guidelines for the development of AI-based medical devices [11]. Thus, new explainable techniques and architectures need to be developed for DNNs to become a trustworthy, transparent, and accountable medical device for clinical decision support. The current project will develop an explainable AI-based telemedical system, called DeepInMotion, able to discover new movement biomarkers of paramount importance for a feasible CP detection and treatment guidance tools. The discovering and verification of new movement biomarkers will lead to a deeper understanding of neurophysiological development of CP and other motor disabilities during childhood.

The main objective of the project is to develop the DeepInMotion system and clinical service implementation to discover movement biomarkers for early detection of CP in infants (see Figure 1 and 2 below).

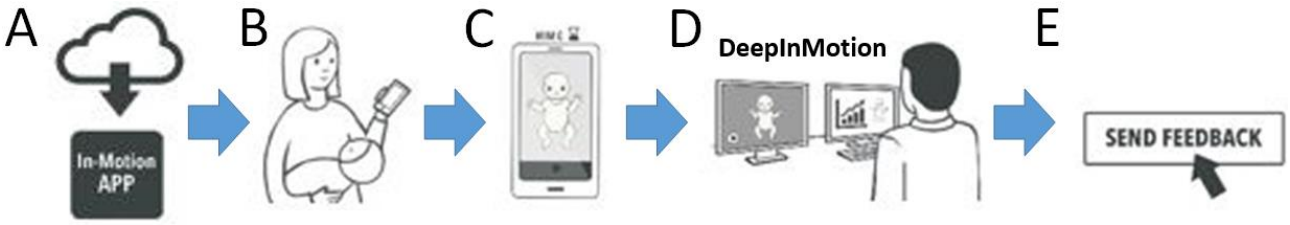


Figure 1: The DeepInMotion clinical service implementation. Eva Johnsen gives birth to her first child, Adam, after 27 weeks of pregnancy. Adam has a small brain hemorrhage during the neonatal period. After discharge from hospital Adam has follow-up scheduled at the hospital follow-up clinic at 6- and 12-months post-term age. Due to the increased risk of motor disability related to preterm birth and the brain hemorrhage, his parents are informed at time of discharge that they will have the opportunity to do detection of CP and severity of disease when Adam is 12-18 weeks post term age. The hospital provides them with an InMotion smart phone app (**A** and **B**) to use at home to take a 5 minutes video recording of Adam while he is placed supine moving spontaneously (**B** and **C**). When Adam reaches 13 weeks post-term age, the InMotion smart phone app reminds Eva (**B**) to perform the video recording following developed standards [12] and upload it to the hospital. An assessment of CP risk and severity of disease is performed based on the DeepInMotion system (see Figure 2 below) by the follow-up clinic (**D**). The clinical dashboard of DeepInMotion shows that Adam has a high proportion of risk-related movements (i.e., biomarkers) identified by DeepInMotion and, thus, a high risk of CP with symptoms indicating that Adam will not be able to walk independently without a mobility device (**D**). The results from the DeepInMotion system is communicated to the pediatrician/physiotherapist with guidelines for clinical actions (**E**). Simultaneously, Eva receives a message on her InMotion smart phone app that Adams' risk assessment is completed together with an appointment at the follow-up clinic. During that appointment Eva, the local physiotherapist, and the pediatrician at the hospital follow-up clinic discuss the implications of the results from the DeepInMotion system for Adam and the family, and a special early intervention program for Adam is initiated. Due to the early targeted intervention program for the gait dysfunctions, Adam can optimize level of activity without a mobility device and increase social contact with friends and family.

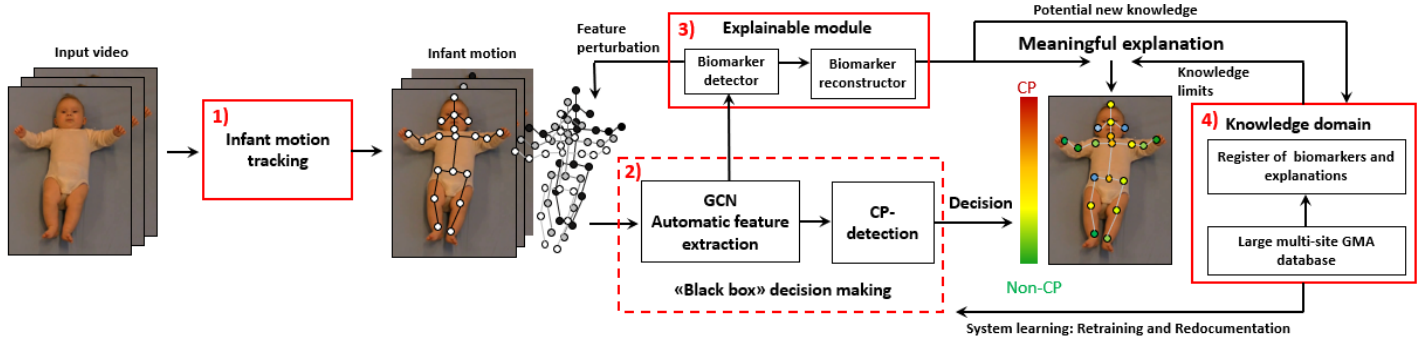
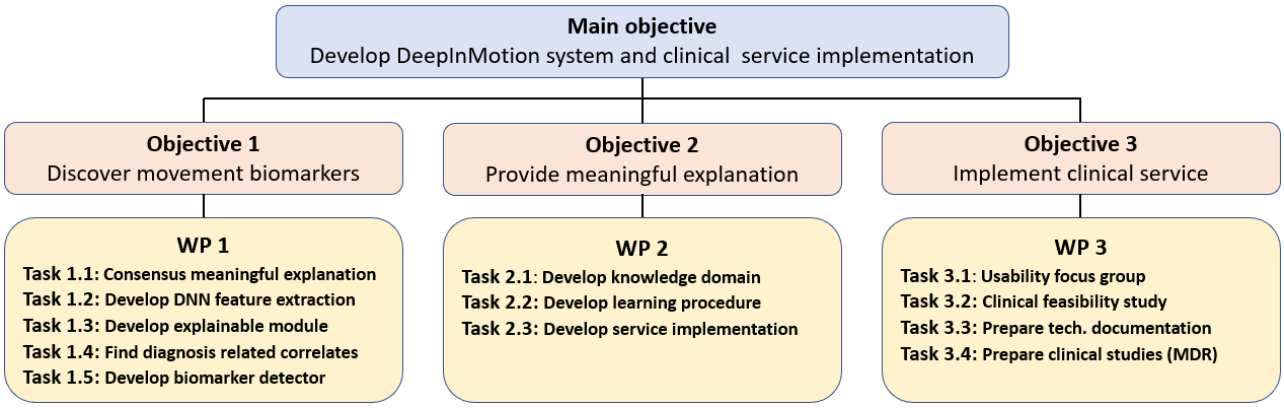


Figure 2: The DeepInMotion system will contain four components: **1)** A DNN for *infant motion tracker* of an input video will identify and track the position of the infant joint centers and body landmarks [13]. **2)** A *DNN for movement feature extraction and CP-detection* will obtain a risk score for development of CP later in childhood. However, the DNN does not provide any explanation for the obtained decision (i.e., decision score). **3)** An *explainable module* with a biomarker detector and reconstructor will provide a decision score (see color code) to each infant joint center for each frame in the input video providing an explanation of the DNN's decision to the end-users. The joint center positions/velocities are perturbed in multiple infant movement simulation (i.e., *feature perturbation*) to find the confidence range of the DNN's decision score. **4)** A *structured knowledge domain* will contain a register of established biomarkers with explanations from a large multi-site GMA data base containing more than 1400 videos of high-risk infants around the globe (Norway, India, Belgium, China, Turkey, US, UK). The large multi-site GMA data base also contain register of severity of CP, CP-subtype, other motor disabilities, and neurophysiological correlates obtained by neuroimaging. The individual decision with biomarkers and explanation will be compared to information contained in the knowledge domain to establish the decision knowledge limit of DeepInMotion which is a part of a meaningful explanation to the end-user. The increase in size and content of the knowledge domain will drive system retraining and learning making DeepInMotion an adaptive and autonomous AI devices for clinical decision support. *A written consent is provided by the infant's parents for the use of the image in this figure*.

1.2 Research aims and hypotheses, theoretical approach and methodology

The project will have the following three work packages (WPs) with related hypothesis, tasks, deliverables, and milestones (see Figure below and GANTT chart in Section 3 for timelines).



WP 1: Discovering new biomarkers of infant movements for early detection of CP

Hypothesis 1.1: The DeepInMotion will detect CP with improved accuracy compared state-of the-art GMA and neonatal imaging [3].
Hypothesis 1.2: The movement features automatically identified by the DeepInMotion provides robust and explainable biomarkers for early detection of CP.

Task 1.1 Consensus for clinical meaningful explanations of DNNs: A Delphi consensus study for definitions of relevant criteria for meaningful explanations and knowledge limits of DNNs will be conducted [8]. These criteria will consider different aspects of DNNs explanations including user benefit, societal acceptance, compliance with laws and regulations of medical devices, and system development [8, 11, 14]. The criteria will also consider the time requirement and level of detail of the explanations for different end-users of the DNNs. These criteria will be used to design the explainable module (see Box 3 in Fig. 2 above and Task 1.3 below) in the DeepInMotion system and the clinical dashboard of it's service implementation (see Fig. 1D above and Task 2.3 in WP2).

Task 1.2 Development of movement feature extractor: The joint centre displacements in the video of the infant spontaneous movements will be assessed by our newly developed state-of-the-art infant motion tracker [13] (see Box 1 in Fig. 2 above). The infant's joint centres and segments will be nodes and edges, respectively, in a graph of the infant's skeletal compositions. This sequence of graphs will be the input of a graph convolutional neural network (GCN) developed through a systematic architectural search based on current state-of-the-art GCNs baseline [15] (see Box 2 in Fig. 2 above). The GCN will be developed using the multi-site GMA database and the deep learning cluster at the Department of Computer Science at NTNU. The database will be divided into a training, validation, and test set. The model with the highest performance on the validation set of the database will be externally validated on the test set and compared with the performance of GMA and neonatal imaging [3] by comparison of sensitivity, specificity, positive and negative predictive value, and area under ROC curve (Hypothesis 1.1). The automatic aggregation of movement features within the GCN will be the input of the explainable module in Task 1.3 and Task 1.4 below to verify that these movement features is biomarkers for early detection of CP.

Task 1.3 Development of explainable module for verification of new biomarkers: The explainable module will consist of two components (see Box 3 in Fig. 1 above); **1a)** a movement biomarker detector will assigning a decision score for each infant's joint centre for each video frame that will localize of the CP-risk related movements. Modifications of localization methods, like gradient class activation mapping, shapley additive explanations, and testing with concept activation vectors, will be developed and compared to evaluate the localization consistency of potential biomarkers [16, 17, 18]. **1b)** Perturbation-based detectors will induce small random perturbation of the infant's joint positions and velocities and investigate how these perturbation affects the decision score of the localization methods above. The random perturbation will be small joint flexion/extension/adduction/abduction in the infant skeletal graph, considering the anatomical constraints of infant joint's range-of-motion, and will work as a simulated input graph to the GCN. A series of simulated input will create a confidence interval for the DeepInMotion decision score in terms of joint positions and velocities providing a quantitative description of the biomarker. **2)** A constructor of the biomarker will reconstruct the infant movement kinematics (i.e., joint positions and velocities) of the time periods in the video where CP-risk related movements are localized according to 1a) and 1b) above. Decoding of the GCNs feature extractor and weighting of the joint centre position and velocities by the decision score will reconstruct the kinematics of CP-risk related movements. The reconstructed kinematics will be used as input in a structural

causal model providing a statistical significance for the movement kinematics being a new biomarker for early detection of CP [19] (Hypothesis 1.2). The criteria obtained in Task 1.1 will be used to design the methods within the explainable module and how the results are visualized in a clinical dashboard of the service implementation developed in Task 2.3 of WP2 below.

Task 1.4 Finding diagnosis related correlates of movement biomarkers: The localization of the movement decision score in Task 1.3 will be compared with diagnosis related correlates. CP-subtype (unilateral/bilateral) will be correlated with the spatial asymmetry in the decision score whereas the temporal frequency of CP-decision score will be compared with the severity of disease (i.e., general motor function classification scale, GMFCS). A specification of Hypothesis 1.2 will be that laterality of the infant movements with CP-decision score is related to affected body-side of unilateral CP and less mobility problems (i.e, GMFCS level I-III) and more bilateral distributed infant movements with CP-decision score will be related to more severe non-ambulatory CP function (i.e., GMFCS level IV and V).

Task 1.5 Automatic detection of new movement biomarkers: The temporal appearance of the CP-decision score in Task 1.3 are used as annotation of the multi-site GMA database for retraining of the GCN to detect CP-related biomarkers instead of the CP outcome. The retraining of the GCN will incorporate error correction of the clinical end-users and updates of the structured knowledge domain developed in WP2 below.

Deliverable (D1): DeepInMotion system for discovering new movement biomarkers for early detection of CP in high-risk infants.

Milestones: (**M1.1**) Criteria of clinical meaningful explanations for the DeepInMotion system. (**M1.2**) Performance test of the optimized version of the GCN for early detection of CP. (**M1.3**) Verification of the explanations of new biomarkers provided by an explainable module. (**M1.4**) The final version of DeepInMotion developed (see GANTT in Section 3.2 below for timelines).

Delivery risk: There is a potential risk for inferior motion tracking of the infant spontaneous movements in the video recordings. To cope with this challenge, previous projects in our group has developed and validated an infant motion tracker which outperforms all current human pose estimation methods [13]. This motion tracker has also been implemented in an easy-to-use software package ready to be shared with clinical partners. There is a potential risk of small sample sizes in the training of the DNN in Task 1.2 and 1.3 even with the large multi-site GMA data base. However, in contrast to conventional statistical prediction models, the DNNs will use pretrained back-ends of large data base of general human movements (i.e., transfer learning) and data augmentation which will make the DNN in Task 1.2 and 1.3 more robust even for small sample sizes. There is also a potential risk of delivering inferior performing DNN models in Task 1.2 and 1.3. The PI of the project has developed and validated a preliminary machine learning model detecting CP within a large cohort of high-risk infants with 92% sensitivity and 81% specificity which is comparable to expert human decisions [10]. However, the model does not provide meaningful explanation for its' decisions to end-users and, thus, is currently unsuitable as a medical device in a clinical service implementation. Nevertheless, the results indicate that the video data contains the information necessary for a highly precise detection of CP by the DeepInMotion. Thus, WP1 is a "low risk and high gain" part of the project.

WP 2: Providing meaningful explanation to the clinical end-users and stakeholders

Hypothesis 2.1: DeepInMotion clinical service implementation will provide meaningful and accurate explanations of its decisions to the clinical end-users and stakeholders.
Hypothesis 2.2: The learning procedure of DeepInMotion will be conform with future effective and secure EU-MDR framework for adaptive and autonomous AI devices for clinical decision support.

Task 2.1 Development of a structured knowledge domain: A structured knowledge domain will be developed with four elements: 1) Multi-site GMA data base of annotated video recordings of the infant spontaneous movements. The video recordings will be annotated according to known motor phenotypes of clinical GMA like fidgety, monotonous, stiff, cramped, and synchronized infant movements [20] and new phenotypes (i.e. movement biomarkers) discovered by the DeepInMotion system. 2) A case-based reasoning system (i.e., k-nearest neighbours) comparing the individual infant movement features with all infant motor phenotypes contained in the Multi-site GMA database. The system will provide an overview of all cases with similar motor phenotypes as the individual infant. 3) Textual analysis of multi-site GMA database, CP registers, electronic patient journals, and clinical literature to provide an evidence-based explanation to each of all the motor phenotypes contained in the database. 4) A gateway of potential new knowledge from the explainable module designed with a human gatekeeper interface where the clinical end-user can error correct

the knowledge input to the GMA database. The structured knowledge domain will communicate the knowledge limits of the DeepInMotion's decisions in the explanations to the clinical end-user making the system more trustworthy as a medical device. The elements of the structural knowledge domain will be developed in close collaborations with the end-users by the usability focus group in Task 3.1 of WP3 below.

Task 2.2 Development of automatic and adaptive learning procedure: As the size and content of the structure knowledge domain increases, the DeepInMotion system will automatically adapt and improve by retraining and learning. The retraining and learning procedure will be developed with four important elements: **1)** A controller network searching for a GCN baseline architecture for optimal performance, computer efficiency and explanation accuracy for early detection of CP [21]. **2)** An efficient grid search procedure to obtain the optimal scaling of depth and width of the GCN baseline architecture to continually match the capacity of the GCN with the size and content of the structured knowledge domain [22]. **3)** A retraining schedule of DeepInMotion initialized according to the change in size and content of the structured knowledge domain incorporating 1) and 2) above. **4)** An automatic documentation procedure of technical specification and performance obtained by 1) and 2) according to EU-MDR guidelines that potentially can be stored directly in European database for medical devices (EUDAMED) for use by notified bodies [10, 14]. Task 2.2 will be performed in close collaboration with Task 3.3 in WP3 below for the learning procedure of DeepInMotion system to become conform with future effective and secure EU-MDR framework for adaptive and autonomous AI devices (Hypothesis 2.2).

Task 2.3 Development of clinical service implementation: The clinical service implementation will be developed with three components: **1)** A smartphone app for video recording of the infant spontaneous movements to be use by health care personnel and infant's parents. The infant's parents will also receive feedback from the follow-up clinic in the same app (see Fig. 1A). **2)** A digital telemedical back-end providing a secure transfer and storage of the infant video recording according to the ICT system at St Olavs Hospital. **3)** A clinical dashboard with graphical user interface (GUI) for the clinical end-user at follow-up clinic (see Fig. 1D). Component 1) and 2) have already been developed for a reginal study in our group for GMA assessment. The development of the clinical dashboard will be done in close collaboration with the development of the explainable module (Task 1.3) and the structured knowledge domain (Task 2.1). The DeepInMotion clinical service implementation will be developed in close collaboration with the health care personnel by a user-centred design with two iterations delivering a low- and high-fidelity version, respectively. Each iteration ends with a usability evaluation by a focus group of clinicians (Task 3.1 in WP3 below).

Deliverable (D2): DeepInMotion clinical service implementation (see Fig. 1 above)

Milestones: (**M2.1**) Development of the structured knowledge domain completed. **(M2.2)** Development of DeepInMotion learning procedure completed. (**M2.3**) High-fidelity version of the clinical service implementation of DeepInMotion completed and feasibility tested.

Delivery risk: The clinical service implementation of the DeepInMotion is a low risk part of the project because it will utilize a prototype of a digital service implementation already developed in another project (GMA service implementation project, REK ID 62240, DPIA 197) ran by our group. The system development of the structural knowledge domain in Task 2.1 will also involve 4-8 master students at Department of Computer Science at NTNU during the first two year of the project.

WP 3: Clinical service implementation of DeepInMotion as a medical device

Hypothesis 3.1: The DeepInMotion clinical service implementation is more feasible in specialist health care services compared with GMA.
Hypothesis 3.2: A Declaration of Conformity (DoC) for DeepInMotion as a Medical Device can be approved by EU-MDR and local licence authorities of developmental countries.

Task 3.1 Usability focus group, containing 5 clinicians at St. Olavs Hospital and Ålesund Hospital and 5 parents (i.e., primary care givers) will participate in qualitative interviews and answer system usability questionnaires assessing the usability and user experience of DeepInMotion clinical service implementation. The usability of the DeepInMotion will be assessed through two iterations during the life-time of the project including the 1) low-fidelity and 2) high-fidelity version of DeepInMotion system and its clincal service implementation developed in WP1 and WP2. Three representatives of the focus group will also be members of the project advisory board.

Task 3.2 Clinical feasibility study will contain two parts: **1)** A small system usability study of the final version of DeepInMotion clinical service implementation at St Olavs Hospital, Levanger Hospital, and Ålesund Hospitals in the Central Norway Regional Health Authority (CNRHA) where 5 physiotherapist and 3 pediatricians will use DeepInMotion on about 30 videos from high-risk infants in a six-month period. **2)** A blinded feasibility study to compare the explanation of a) clinical GMA and b) DeepInMotion service implementation. Videos from 100 high-risk infants collected in an ongoing digital GMA service implementation study in the CNRHA between 2019 and 2022 (REK ID 62240, DPIA 197) will be used to compare the explanation accuracy and meaningfulness of a) and b) assessed by a group of 10 physiotherapists and paediatricians using questionnaire (Hypothesis 3.1). The usability and feasibility study will adhere to the standards of EU-MDR and be a pilot for preparation of clinical follow-up for DoC (see Task 3.4 below). The clinical feasibility study will be included European database for medical devices (EUDAMED) if included in the evaluation by EU-MDR notified body.

Task 3.3 Preparing pre-market technical documentation for DoC: The development and documentation of components illustrated in Fig. 2 will follow the audit guidelines for MDR notified body for AI-based medical software and EU-MDR 2017/745 [11,14]. A joint GitHub code repository across all WPs will be established in the beginning of the project following the audit guidelines and the consensus established in Task 1.1 in WP1. The code repository will also contain a list of third-party software dependencies for the algorithms within DeepInMotion system (i.e., SOUP in IEC 62304) ready to be filed to the quality management system and EUDAMED. The intended use of DeepInMotion system will be fine-tuned and classification according to EU-MDR and MDCG 2019-11 will be established. The technical documentation will adhere to a full list of medical device directives of MDR Annex II that apply to the appropriate classification of DeepInMotion system. An auto-documentation procedure for periodic safety update report to EUDAMED will be developed for DNN learning and adjustments which adhere to the list of directives for post-market surveillance of the DeepInMotion system (see also Task 2.2 in WP2 above). In special cases where the EU-MDR documentations do not follow the local directives for licence approval in developmental countries, the documentation will be adapted under supervision of the audit guidelines of the local licence authorities.

Task 3.4 Preparing clinical studies for DoC: A dated list of all standards and common specifications used to evaluate DeepInMotion according to EU-MDR will be created. In this list, the general safety and performance requirements of MDR Annex I to intended use will be of central importance. Thus, the preparation of clinical studies will contain two parts: **1)** external validation of the performance of the DeepInMotion for early detection of CP and **2)** external clinical feasibility tests of a product version of DeepInMotion. Both the protocol of the external validation and feasibility studies will follow the ISO 13485 standard needed for quality management system of medical devices to ensure compliance with the EU-MDR [23]. Collaboration between clinical and medical device manufacturers in advisory board (i.e., industrial partners) will be in accordance to these standards and other requirements from the EU-MDR. The industrial partners will exploit the IP of the project through license and patents arrange by the NTNU Technological Transfer Office (TTO) and will be responsible for a product version of clinical service implementation of DeepInMotion including both the front-end and back-end illustrated in Fig. 1 and 2 above, respectively. EU-MDR requirements for product specifications, unique device identification (UDI), post-market surveillance, product-user guides, etc will be responsibilities of the product manufacture and, thus, beyond the scope of the current project.

Deliverable (D3): Pre-market technical documentation and clinical follow-up preparations of DeepInMotion necessary to meet the EU-MDR requirements for a DoC of medical devices.

Milestones: (**M3.1**) All technical documentation for DeepInMotion required by EU-MDR is prepared. (**M3.2**) All protocols for clinical-follow up required by EU-MDR is prepared.

Delivery risk: The performance, explanation accuracy, and knowledge domain of AI-based software as medical device (SaMD), like DeepInMotion, will constantly improve during clinical use by system retraining and learning. Thus, an MDR for AI-based SaMD requires a new total product lifecycle regulatory approach to provide an effective and secure regulatory framework for adaptive and autonomous AI devices. Currently, the regulatory for implementing AI systems as medical devices are still under development and there is a risk that regulatory framework may change during the project period. To minimize this risk, the project will follow preliminary FDA guidelines for developing AI-based SaMD [11] and continuously collaborate with the creators of the audit guidelines for EU-MDR notify bodies to be inform on upcoming changes. In addition, the documentation procedures developed in Task 2.2 (WP2) and Task 3.3 have a high potential to be effective and secure and adhere to future product lifecycle regulatory for AI-based SaMD.

Interdisciplinary approach: The present project has a unique interdisciplinary collaboration between clinicians, human movement scientists, computer scientists, and medical device manufacturers in all WPs to obtain a new concept for clinical movement analysis for improved disease detection and treatment. Clinical scientists will evaluate the clinical feasibility of the DeepInMotion system (WP2 and WP3) whereas the computer scientists in collaboration with human movement scientists will develop DNNs of movement feature extraction, biomarker detection, diagnosis-prediction, and end-user explanation (WP1 and WP2). The interdisciplinary approach will be facilitated by a user-centered interative development procedure where results of the usability tests of the low-fidelity version of the DeepInMotion (WP2) will be used to improve subsequent high-fidelity versions of the system (WP1 and WP2). Two medical device manufacturers (Ascom and Distributed Medical) will be members of the project advisory board as industrial partners to exploit project IPs and collaborate on the preparation of technical documentation and clinical follow-up necessary to meet the EU MDR requirements for a DoC for the DeepInMotion system (WP3).

Ethical perspectives: The infants' families will give consent for the video data to be processed by the DeepInMotion according to European General Data Protection Regulation (GDPR). A risk and vulnerability (ROS) analysis of the project, including data storage, transfer, and implementation, will be performed by NTNU. Furthermore, the use of AI on person sensitive data will be in accordance with GDPR and national regulation and AI methodology able to re-identify anonymous data will not be used [24]. No sensitive medical information will be communicated outside the secured ICT-system (PAS) of St. Olav hospital. Validation-, usability-, and feasibility studies that are not approved will be presented to the Regional Committees for Medical and Health Research Ethics and documented in EUDAMED as part of the EU-MDR. An essential aim of the project is to promote equality in access to health services by developing services with explainable AI techniques which can be implemented in high-, mid- and low-resource settings world-wide.

Gender issues (Recruitment of women, gender balance and gender perspectives): The global multi-site GMA database have an equal gender distribution which is important for the validation of the DeepInMotion system. The research groups behind the project consist of both genders.

Use of stakeholders/user knowledge: End-user knowledge on clinical feasibility will be utilized in development of DeepInMotion system and its service implementation in WP1 and WP2 (i.e., paediatricians, physiotherapists, other health care personnel, and family of the infant/children). The knowledge will be obtained in usability focus groups in Task 3.1 in WP3 and through members of the advisory board containing both end-users, medical device manufacturers, and creators of audit guidelines for EU-MDR notified bodies.

1.3 Novelty and ambition

The project will develop new knowledge beyond the current state-of-the-art by the creation of DeepInMotion system addressing the fundamental challenges in clinical movement analysis and machine learning-based disease detection. In WP1, confirmation of Hypothesis 1.1 will provide a paradigm shift in clinical movement analysis by the development of the first system ever to automatically extract new movement features which is directly related to CP as a diagnostic outcome (Task 1.2). The novel graph convolutional network (GCN) architectural design obtained by Task 1.2 and modified and scaled in Task 2.2 will be the first GCN design ever to be developed for clinical movement analysis and decision support. The GCN will contain advancements in network layer architecture and optimization procedures that will improve precision, computer efficiency, and explanation accuracy compared to former state-of-the-art GCN [15]. Confirmation of Hypothesis 1.2 will provide each extracted feature by the GCN with a clinical meaningful explanation defining the extracted movement features as biomarkers for the diagnostic outcome (Task 1.1, 1.3, 1.4, and 1.5). By the development of new algorithms for spatiotemporal localization of GCN decisions and confidence intervals in terms of infant joint position and velocities, the explainable module will be the first ever to localize and reconstruct the infant movement biomarkers to be displayed to clinicians in a meaningful way. Thus, confirmation of Hypothesis 1.1 and 1.2 together will provide the world's first autonome AI-based detection of human movement biomarkers to substitute todays subjective and qualitative movement assessments [3, 5]. In WP2, the confirmation of Hypothesis 2.1 will provide a stronger establishment of meaningful, trustworthy, and accurate explanations of movement features by the construction of a structured knowledge domain. The structured knowledge domain will contain a case-based reasoning system linking the per-decision movement features to motor phenotypes (i.e., biomarkers) of previous patient cases and their prognosis and progression of disease (Task 2.1). The textual and graphical visualizations in the clinical dashboard (Task 2.3) will be adapted to the end-user and stake holders through user-centred design to provide meaningful explanations. In WP3, the confirmation of

Hypothesis 3.1 will establish a XAI-based clinical service that are an improvement of current GMA service for early detection of CP and, thus, promote well-being for all children reducing inequality of health care within and among countries (UN sustainable development goal 3 and 10). Confirmation of Hypothesis 3.2 will be an advancement in knowledge on documentation procedures for efficient transfer of IP obtained from conceptual research within computer and human movement science to medical device manufacturers. The deliverable of Task 3.3 and Task 2.2 of WP2 will provide a novel automatic DNN architectural update procedure according to continuous change in the knowledge domain of infant spontaneous movements and motor phenotypes. This novel procedure will be the first ever to provide an automatic documentation of product life-cycle changes in AI-based medical software system according to EU-MDR guidelines and, thus, will be an advancement in secure and efficient post-market surveillance of AI-based medical software devices [11, 14]. In summary, the DeepInMotion will bridge the technical-clinical knowledge gap improving acceptance and translation of AI-based medical software into clinical practice.

2. Impact

2.1 Potential impact of the proposed research

The expected results of WP1 will provide a paradigm shift in clinical movement analysis by the development of explainable machine learning algorithms building procedures to discover biomarkers of disease outcome and their relation to neuroanatomical abnormalities. The expected results of WP2 will provide clinical service implementation of the new explainable AI techniques for early detection of CP independent of geographic affiliation. The expected results of WP3 will be a pre-market documentation and verification of the new explainable AI techniques to facilitate EU-MDR approval for AI-based medical software. Together, the expected results of WPs will be a low-cost digital AI-based health care service reducing inequalities within and among countries and providing equal rights for health care services (UN sustainable development goal 3 and 10) by improving help and support to clinicians and primary care givers in a safe and efficient manner [25, 26, 27].

2.2 Measures for communication and exploitation

The exploitable results, target audience, tools and channels, and expected impact is summarized in Table below. The target audience will be researchers within computer science, medical technology, movement science, and medicine (R&D), health care personnel (HCP; paediatricians and physiotherapists), primary care givers (PCG; infants' family), and medical device manufacturers and EU license authorities (MDM). The measure of communication will be the following tools and channels adapted to the target audience (see Table below): **1)** Project website (PW) and social media (SM) including blogs, Facebook, and Twitter. **2)** Open source, peer-reviewed scientific journals, Research Gate, and conferences (SJC) and GitHub algorithm repository (GHR) (see also '*Dissemination plan*' in *electronic grant application form* for further details). **3)** Annual seminar, workshops, and showcases (SWS). Table below summarize the tools and channels for each target audience together with the expected impact for each WP in the project. The exploitable results of WP1 and WP2 will be declared as an innovation (DOFI) and an intellectual assets management plan and patents for the DOFI will be developed, in collaboration with technological transfer office at NTNU, for potential transfer of IP to medical device manufacturers represented in the project advisory board.

WP	Name	Exploitable results	Audience	Tools and channels	Expected impact
WP 1	**Discovering new movement biomarkers**	Explainable DNN architectures for clinical movement analysis (Declaration of Innovation)	R&D	SJC GHR	Facilitate research on explainable DNNs for clinical movement analysis
			MDM	PW SM	Facilitate development of explainable AI-based medical software
WP 2	**Providing meaningful explanation to end-users and stakeholders**	Clinical service implementation of explainable DNNs (Declaration of Innovation)	R&D HCP PCG	SJC GHR SWS, SM	Equal opportunity for early detection of CP independent of geographic affiliations
			MDM PCG	PW SM	Low-cost AI-based health care services to reduce inequalities within and among countries
WP 3	**Implementing clinical service**	Efficient documentation procedure for AI-based medical software	R&D MDM	SJC GHR SWS	Facilitate EU-MDR and US-FDA approval for AI-based medical software

Abbreviations Audience: R&D = Research and development stakeholders, HCP = Health care personnel, MDM = Medical device manufacturers and EU/US license authorities, PCG = Primary care givers

Abbreviations Tools and channels: SJC = Scientific Journals & Conferances, PW = Project website, SM = social media, GHR = Git Hub algorithm repository, SWM = Seminar, workshops, and showcases

Dissemination plan and **budget**: See **electronic grant application form** for further details

3. Implementation

3.1 Project manager and project group

Project manager (PI) expertise and experience: The PI of the project will be associate professor **Espen A. F. Ihlen**. Despite his short post PhD graduate period (6 years) he has 40 publications in peer-reviewed international journals (20 publications as first author where 10 is in level 2 journals) and 30 oral presentation (first author and co-author) and 10 posters at international conferences (H-index of 17). The scientific works centres around methodological development within the research field of movement analyses and is consistent with interdisciplinary strategies for NTNU in medical, health and welfare technology. As a PhD student, he coordinated the writing of the project proposal to Norwegian Research Council; A personalized case risk assessment system for promoting independent living (The ADAPT project), which was awarded the grant of 9.3 Mill NOK in 2013. He has contributed to the development of concepts within the European Horizont 2020 Projects «selfBACK» and "PreventIT" which were awarded with grants in 2015. He has been one of the main facilitators of the inter-disciplinary collaboration between Department of Neuromedicine and Human Movement Science (INB), Department of Clinical and Molecular Medicine (IKOM), and Department of Computer Science (IDI) at NTNU and two clinics at St Olavs Hospital (NICU and Clinic of Clinical Services) focusing on clinical implementation of new machine learning techniques. In the years to come he will establish the inter-disciplinary research group, DeepMotion, on clinical movement analysis to develop new and improved analysis for diagnosis, recommendation of treatment and prognosis for motor disabilities like CP.

Research group expertise and experience: The project is a start-up of the DeepMotion research group which is an interdisciplinary collaboration between several departments of human movement science, computer sciences and clinical sciences at NTNU and St Olavs hospital in Trondheim, Ålesund Hospital, and world leading universities and hospitals within their scientific fields. The allocation of tasks to each project team member is represented in Section 3.2 below: *Human movement science*: Department of Neuromedicine and Human Movement Science (INB) at NTNU, represented by associate professor **Espen A.F. Ihlen (PI)**, will be the owner of the project and covers research on movement science and movement disorders with a special focus on CP. In the national research evaluation for medicine and biology in 2012, the research groups at INB was rated as very good to excellent [28]. INB runs projects spanning from mathematical modelling and laboratory experiments on the understanding of movement problems, to method development and evaluation studies, particularly using movement analysis, and clinical studies on prevention, treatment and rehabilitation for different patient populations. *Computer science*: The Department of Computer Science host the Norwegian Open AI-lab, represented by professor **Heri Ramampiaro** and professor **Helge Langseth**, which covers innovative research on artificial intelligence, machine learning, and big data analytics with application to diagnosis and treatment of CP patients as one of their main focus areas. Distributed, Embedded and Intelligent Systems (DEIS) at Department of Computer Science at University of Ålborg, represented by associate professor **Thomas Dyhre Nielsen**, are experts in probabilistic network architecture and machine learning. *Clinical science*: Professor **Ragnhild Støen** are head the Children Clinic and the level III Neonatal intensive care unit at St Olavs hospital caring for preterm and other sick infants, many of whom are at increased risk of adverse development. Associate professor and pediatricians **Beate Horsberg Eriksen** represents the Children Clinic at Ålesund Hospital and are head of the "Norsk perinantalmedisinsk forening". Senior Researcher **Lars Adde** represent the Clinic of Clinical Services at St Olavs hospital responsible for follow-up of high-risk infants and children with CP and are PI of a project on implementation of a digital service implementation of GMA at several hospitals in mid-Norway (REK ID 62240, DPIA 197). The clinics cover research on neonatal care, advanced MRI techniques, GMA and neurodevelopmental follow-up of infants with risk of motor dysfunctions and coordinates the multisite GMA project with the world largest data base of standardized video recordings of infant movement repertoire. NICU follow-up program at the Nationwide Children hospitals in US, represented by Director and Professor **Natalie J. Maitre,** is world's leading on the development of quantitative methods for neural and motor function assessment for early identification of children with high-risk of motor disorders. The Department of Neuroscience at University of Copenhagen in Denmark, represented by Professor **Jens Bo Nielsen**, is leading on research on basic principles and mechanisms of motor learning, the neuroplastic changes in the central nervous system (CNS) and the relationship between neuroplasticity and behavioural changes in infants and children.

3.2 Project organisation and management

Project plan: The GANTT chart below summarizes the timelines for all project tasks, milestones, and deliverables (D1, D2, and D3) defined in Section 1.3 above. The project starts in October 2021. The positions responsible for the different tasks are indicated by coloured bars in the GANTT chart below. A 35% associate professor (PI) position funded by NTNU will coordinate all tasks in the project (green bars) and be responsible for the preparation of the technical documentation to the EU MDR (Task 3.3). A 100% 3-year PhD position funded by NFR will perform all tasks in WP1 except Task 1.1 (yellow bars). A 100% 3-year PhD position funded by NFR will perform all tasks in WP2 (orange bars). A 30% 3-year engineering position funded by NFR will develop DeepInMotion clinical service implementation (Task 2.3) into PAS system of St Olavs hospital in collaboration with ICT services of the hospital (HEMIT) and medical device manufacturers in the advisory board. A 3.5-year 30% clinical research coordination position at St Olavs Hosptital funded by NFR will be WP3 leader and responsible for the consensus process (Task 1.1), usability focus groups (Task 3.1), clinical feasibility studies (Task 3.2 and 3.4) and data management of the multi-site GMA data base (purple bars). A total of 12 person months (physiotherapists, paediatricians) will be dedicated to St Olavs and Ålesund hospital for data and study management costs of Task 3.1 and 3.2. The PhD fellowships will have at least three supervisors, where at least one of the supervisors is a senior researcher and have conducted the NTNU course for PhD supervision and where one is a clinician. Project-wide milestones related to WP 1-3 above are summarized as vertical lines in the GANTT chart. See the electronic grant application form for more information on the project wide activities and milestones. Carbon prints due to travels will be kept to a minimum by digital meetings and combining conferences with workshops with the international partners.

Tasks		2021				2022				2023				2024				2025			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
WP 1	**Discovering new biomarkers**																				
1.1	Consensus for meaningful explain.																				
1.2	Develop DNN feature extraction																				
1.3	Develop explainable module																				
1.4	Find diagnosis related correlates																	D1			
1.5	Develop biomarker detector																	D1			
WP 2	**Providing meaningful explanation**																				
2.1	Develop knowledge domain															D2					
2.2	Develop learning procedure																	D2			
2.3	Develop service implementation																	D2			
WP 3	**Implementing clinical service**																				
3.1	Usability focus group																				
3.2	Clinical feasibility study																			D3	
3.3	Preparing technical documentation																			D3	
3.4	Preparing clinical studies MDR																			D3	
	Principle Investigator																				
	Milestones			Start NFR fund		M1.1				M1.2				M1.3		M2.1		M2.2 M2.3 M1.4		M3.1 M3.2 M3.3	End NFR

Allocation of tasks to project team members and collaboration partners: The following table defines the allocation of task to project team members and collaboration partners:

Tasks	Name	Team members	Collab. partners	Clinics
WP1	**Discovering new biomarkers**			
1.1	Consensus process	EAFI, HR, LA	JBN, NJM	SOH
1.2 to 1.5	Supervisors PhD student	EAFI, HR, HL, LA	TDN	
WP 2	**Provide meaningful explanation**			
2.1 to 2.3	Supervisors PhD student	EAFI, HR, HL, LA	TDN	
WP 3	**Implementing clinical service**			
3.1	Usability focus group	LA, RS, BHE	JBN, NJM	SOH, AH
3.2	Clinical feasibility study	LA, RS, BHE	JBN, NJM	SOH, AH
3.3	Prepare technical documentation	EAFI, HR, HL	TDN	
3.4	Prepare clinical studies	LA, RS, BHE	JBN, NJM	SOH, AH
All Tasks	Project manager	EAFI		
All Tasks	Data managers: GMA data base	LA, RS		SOH
All Tasks	DNN computer cluster manager	HR, HL		

Abbreviations: EAFI = Espen A.F. Ihlen, HR = Heri Ramampiaro, LA = Lars Adde, HL = Helge Langseth, RS = Ragnhild Støen, BHE = Beate Horsberg Eriksen, NJM = Natalie J. Maitre, JBN = Jens Bo Nielsen, TDN = Thomas Dyhre Nielsen, SOH = St Olavs Hospital, AH = Ålesund Hospital

Organization and management structure and user/stakeholders involvement: The management structure will consist of an interdisciplinary steering group containing 2 members form NTNU (EAFI, HR), 2 members from St. Olavs Hospital (RS, LA) and 1 member from Ålesund Hospital (BHE) and a project consortium including additional 4 project partners (HL, TDN, JBN, NJM). The project will also contain an advisory board

of end-users and stakeholders of DeepInMotion system including 1-3 representatives of medical device manufacturers, 1-3 representatives from primary care givers and 2 representatives from technological transfer office (TTO) at NTNU. The advisory board will take part in the development of DeepInMotion clinical service implementation (Task 2.3) and pre-market documentation and study design for MDR (Task 3.3 and 3.4). The advisory board will also be involved in arrangement of annual events and showcase of the DeepInMotion clinical service implementation. The internal communication for the team members, collaboration partners, and clinics will be monthly digital meetings and annual physical meeting/workshop. Our research group will have weekly scientific PhD meetings and a long-term strategy for the research as well as for education of the PhD students of the project to become future post-doctoral scholars and master/PhD supervisors.

Research infrastructure (WP1 and WP2): The Norwegian Open AI-Lab will provide the necessary computational infrastructure for this project including a state-of-the-art deep learning cluster containing 10 Tesla V100 and 54 Tesla P100 GPUs. The host department (INB) will also provide access to NextMove core facility which contains a state-of-the-art laboratory for movement analysis including high-speed video, body worn IMUs, and several marker-based 3D motion capture systems (Vicon and Qualisys).

Description of register data for CP detection (WP1 and WP2): The unique multisite GMA data base contains the world's largest database of over 1400 standardized video recordings of infant movement repertoire 10-15 weeks post term age from sites around the globe; Norway (5 sites), India, Belgium, Turkey, China, UK, and US. The infant movement repertoire is obtained by observer rated GMA of the video recordings performed by authorized physiotherapists. Motor function outcomes in the multisite GMA data base is assessed at 18 months to 4 years, and includes CP status, CP subtype and functional level (Gross Motor Function Scale) and motor function in children without CP assessed by the Movement Assessment Battery for Children (M-ABC). The multisite GMA data base also includes MRI and ultrasound data.

References

[1] Blencowe H, et al. (2013) Estimates of neonatal morbidities and disabilities at regional and global levels for 2010: introduction, methods overview, and relevant findings from the Global Burden of Disease study. Pediatr Res 74, 4-16.
[2] Kruse M, et al (2009) Lifetime costs of cerebral palsy. Dev Med Child Neurol 51(8), 622-8.
[3] Novak I, Morgan C, Adde L, et al. (2017) Early, Accurate Diagnosis and Early Intervention in Cerebral Palsy: Advances in Diagnosis and Treatment. JAMA Pediatr 171(9):897-907.
[4] Hadders-Algra M, Gramsbergen A (2007) Discussion on the clinical relevance of activity-dependent plasticity after an insult to the developing brain. Neurosci Biobehav Rev 31(8), 1213-9.
[5] Maitre, N. (2018) Skeptism, cerebral palsy, and the general movement assessment. Dev Med Child Neurol 60(5), 438.
[6] Marcroft, C. et al. (2015). Movement recognition technology as a method of assessing spontaneous general movements in high risk infants. Front. Neurol. https://doi.org/10.3389/fneur.2014.00284
[7] https://aiimpacts.org/time-for-ai-to-cross-the-human-performance-range-in-imagenet-image-classification/ (retrieved 19.01.2021)
[8] Philips et al. (2020) Four principles of explainable artificial intelligence. https://www.nist.gov/publications/four-principles-explainable-artificial-intelligence-draft (Retrived 31.01.2021)
[9] Silva et al. (2021). The future of General Movement Assessment: The role of computer vision and machine learning – A scoping review. Res Dev Disabil, 110, https://doi.org/10.1016/j.ridd.2021.103854
[10] Ihlen et al. (2020). Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: A multi-site cohort study. J Clin Med. 9(1): 5. https://www.mdpi.com/2077-0383/9/1/5/htm
[11] ai-guideline/Guideline-AI-Medical-Devices_EN.md at master · johner-institut/ai-guideline · GitHub (retrived 31.01.2021)
[12] https://youtube/YwdRnl0HocE (retrived 31.01.2021)
[13] Groos, D, Adde, L, Støen, R, Ramampiaro, H, Ihlen, EAF (2020) Towards human performance on automatic motion tracking of infant spontaneous movements. https://arxiv.org/abs/2010.05949
[14] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745 (retrieved 29.12.2020)
[15] Song et al. (2020) Stronger, faster, and more explainable: A graph convolutional baseline for skeleton-based action recognition. https://arxiv.org/pdf/2010.09978v1.pdf
[16] Lundberg et al (2017). A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 4765–4774.
[17] Chattopadhay et al., (2018). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. IEEE WACV2018, 839–847
[18] Kim et al. (2017) Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV). https://arxiv.org/abs/1711.11279
[19] Reimers et al. (2020). Determining the relevance of features for deep neural networks. ECCV2020, 330-346.
[20] Einspieler et al. (2019). Cerebral palsy: Early markers of clinical phenotype and functional outcome. J Clin Med. 8(10): 1616.
[21] Tan, M. et al. (2019) MnasNet: Platform-aware neural architecture search for mobile. CVPR, 2820-2828.
[22] Tan, M., Le, QV. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, PMLR, 97: 6105-6114,
[23] https://www.standard.no/no/Nettbutikk/produktkatalogen/Produktpresentasjon/?ProductID=881541
[24] https://www.datatilsynet.no/globalassets/global/om-personvern/rapporter/rapport-om-ki-og-personvern.pdf
[25] https://www.un.org/sustainabledevelopment/health/
[26] https://helsedirektoratet.no/samhandlingsreformen
[27] https://ehelse.no/publikasjoner/nasjonal-handlingsplan-for-e-helse-2017-2022
[28] Research in Information and Communication Technology in Norway: An evaluation, Research Council of Norway, March 2012