Author/s:
Spittle, AJ;Marschik, PB;Adde, L;Badawi, N;Byrne, R;Bos, AF;Chatelin, A;Coughlan, J;Fedeli, F;Guzzetta, A;Ho, ESL;Johnson, MJ;Kwong, A;McEwan, A;Morgan, C;Mughogho, A;Murray, DM;Orlandi, S;Peyton, C;Prosser, LA;Ritterband-Rosenbaum, A;Tran, T;Zhang, D;Passmore, E

Title:
Towards universal early screening for cerebral palsy: a roadmap for automated General Movements Assessment

Date:
2025-08-01

Citation:
Spittle, A. J., Marschik, P. B., Adde, L., Badawi, N., Byrne, R., Bos, A. F., Chatelin, A., Coughlan, J., Fedeli, F., Guzzetta, A., Ho, E. S. L., Johnson, M. J., Kwong, A., McEwan, A., Morgan, C., Mughogho, A., Murray, D. M., Orlandi, S., Peyton, C. ,... Passmore, E. (2025). Towards universal early screening for cerebral palsy: a roadmap for automated General Movements Assessment. Eclinicalmedicine, 86, pp.103379-. https://doi.org/10.1016/j.eclinm.2025.103379.

Persistent Link:
https://hdl.handle.net/11343/367517

Towards universal early screening for cerebral palsy: a roadmap for automated General Movements Assessment

Alicia J. Spittle,[a,b,c,*] Peter B. Marschik,[d,e,f,g] Lars Adde,[h,i] Nadia Badawi,[j] Rachel Byrne,[k] Arend F. Bos,[l] Alain Chatelin,[m] John Coughlan,[n] Francesca Fedeli,[o] Andrea Guzzetta,[p,q] Edmond S. L. Ho,[r] Michelle J. Johnson,[s,t] Amanda Kwong,[a,b,c] Alistair McEwan,[j,u] Catherine Morgan,[j] Anderson Mughogho,[v] Deirdre M. Murray,[w,x] Silvia Orlandi,[y,z,ai] Colleen Peyton,[aa,ab] Laura A. Prosser,[s,t] Anina Ritterband-Rosenbaum,[ac] Truyen Tran,[ad] Dajie Zhang,[d,f] and Elyse Passmore[ae,af,ag,ah]

[a]Department of Physiotherapy, Melbourne School of Health Science, University of Melbourne, Melbourne, Australia
[b]Clinical Sciences, Murdoch Children's Research Institute, Melbourne, Australia
[c]Department of Physiotherapy, The Royal Women's Hospital, Melbourne, Australia
[d]Department of Child and Adolescent Psychiatry, University Hospital Heidelberg, German Center for Mental Health (DZPG), Heidelberg University, Heidelberg, Germany
[e]Child and Adolescent Psychiatry and Psychotherapy, University Medical Center Göttingen, German Center for Child and Adolescent Health (DZKJ) and Leibniz ScienceCampus Primate Cognition, Göttingen, Germany
[f]Division of Phoniatrics, interdisciplinary Developmental Neuroscience (iDN), Medical University of Graz, Graz, Austria
[g]Center of Neurodevelopmental Disorders (KIND), Center for Psychiatry Research, Department of Women's and Children's Health, Child and Adolescent Psychiatry, Region Stockholm, Karolinska Institutet & Stockholm Health Care Services, Stockholm, Sweden
[h]Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway
[i]Clinic of Rehabilitation, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
[j]Cerebral Palsy Alliance Research Institute, Brain Mind Centre, Discipline of Child and Adolescent Health, The University of Sydney, Sydney, NSW, Australia
[k]Cerebral Palsy Foundation, New York, NY, USA
[l]Division of Neonatology, Beatrix Children's Hospital, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
[m]Fondation Paralysie Cerebrale, Paris, France
[n]International Cerebral Palsy Society, Bath, United Kingdom
[o]Fight the Stroke Foundation, Milano, Italy
[p]Department of Clinical and Experimental Medicine University of Pisa, Italy
[q]Department of Developmental Neuroscience, IRCCS Stella Maris, Pisa, Italy
[r]School of Computing Science, University of Glasgow, Scotland, United Kingdom
[s]Department of Physical Medicine and Rehabilitation, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[t]Rehabilitation Robotics Lab, GRASP, University of Pennsylvania, Philadelphia, PA, USA
[u]School of Biomedical Engineering, University of Sydney, Australia
[v]Department of Rehabilitation Sciences, Kamuzu University of Health Sciences, Blantyre, Malawi
[w]Dept of Paediatrics and Child Health, University College Cork, Cork, Ireland
[x]INFANT Maternal and Child Health Centre, University College Cork, Ireland
[y]Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi" – DEI, University of Bologna, Bologna, Italy
[z]IRCCS Istituto Delle Scienze Neurologiche di Bologna, Bologna, Italy
[aa]Department of Physical Therapy and Human Movement Sciences, Feinberg School of Medicine, Northwestern University, Chicago, USA
[ab]Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, USA
[ac]Elsass Foundation, Gentofte, Denmark
[ad]Applied Artificial Intelligence Institute, Deakin University, Waurn Ponds, Australia
[ae]Developmental Imaging Group, Murdoch Children's Research Institute, Melbourne, Australia
[af]Gait Analysis Laboratory, Royal Children's Hospital, Melbourne, Australia
[ag]Department of Paediatrics, University of Melbourne, Melbourne, Australia
[ah]Department of Biomedical Engineering, University of Melbourne, Melbourne, Australia
[ai]Health Sciences and Technologies, Interdepartmental Center for Industrial Research (CIRI-SDV), University of Bologna, Italy

Summary

Cerebral palsy (CP) is the most common lifelong physical disability, affecting millions globally. Early detection and intervention are crucial for improving outcomes, yet many children are diagnosed late. The General Movements Assessment (GMA) is a highly accurate clinical tool for detecting infants at high probability of CP, but access to

*Corresponding author. Department of Physiotherapy, Melbourne School of Health Science, University of Melbourne, Melbourne, Australia.
E-mail address: aspittle@unimelb.edu.au (A.J. Spittle).

eClinicalMedicine
2025;86: 103379
Published Online xxx
https://doi.org/10.1016/j.eclinm.2025.103379

health professionals trained in the GMA limits its use. Artificial intelligence (AI) has the potential to automate the GMA, increasing accessibility worldwide. We established an interdisciplinary, international consortium for the purpose of developing a roadmap for the ongoing development and implementation of an AI-enabled GMA system for universal CP screening worldwide. The consortium included clinicians (children neurologists, paediatricians, neonatologists, allied health), researchers, engineers, computer scientists, legal experts, and individuals with lived experience, from around the globe (across Africa, Australia, Europe, and North America). The roadmap identifies the following steps and key requirements within: (1) development of standards for AI validation; (2) development of AI-GMA from large and diverse validation sets; (3) development of software tools and clinical pathways; (4) regulatory requisites; and (5) implementation. With the roadmap, AI-enabled screening for CP incorporating state-of-the-art technology can be made possible. Future work will require international collaboration to allow for scaling of data sets, refining automated solutions and translation into practice.

Funding Cerebral Palsy Foundation, Cerebral Palsy Alliance, European Union Born to Get There, the National Health and Medical Research Council.

Keywords: Artificial intelligence; Machine learning; General movements; Early detection; Cerebral palsy

Introduction

Cerebral palsy (CP) is a heterogenous group of movement disorders with multiple causal pathways.[1] CP results from a maldevelopment or injury to the developing brain affecting 1–4 per 1000 live births globally, with higher rates in low- and middle-income countries (LMIC).[2] Early detection and intervention are critical for optimising developmental outcomes and quality of life for children and their families.[3,4] International clinical guidelines recommend diagnosis as early as possible using accurate assessment tools in the first months of life.[4] However, many children worldwide lack access to reliable early screening methods, thus delaying intervention.[5–7]

The General Movements Assessment (GMA) is a non-invasive video-based observational tool that assesses the quality of infant spontaneous movement patterns.[8,9] The absence of a specific general movement pattern called "fidgety movements" (FMs)–small multiplanar movements observed in the wrists, shoulders, hips, ankles, and neck–demonstrates a sensitivity of 95–98% for predicting CP when used at 3–4 months of age, making the GMA the most accurate clinical tool to predict CP.[4,10,11] However, widespread implementation of GMA is limited by the need for training and certification of assessors.[12] Recent advances in artificial intelligence (AI) offer a promising opportunity to address these gaps by automating the GMA,[13–18] allowing the potential for universal screening of CP. Universal screening during critical development windows in early life has the potential to reduce diagnostic inequities—particularly in low-resource settings where access to trained assessors is limited.

Video collection using consumer-grade red-green-blue (RGB) cameras, commonly available in personal devices such as smartphones, tablets, laptops, and webcams, are the most accessible and practical option for universal screening of CP.[19] These devices are not only widely available and easy for use in clinical routine settings but also families and caregivers can use them in home environments, making them suitable for screening at scale.[19–23] For this reason, the majority of automated GMA work has been developed using video recordings from RGB cameras.[14,17,18,24]

The process of AI GMA prediction typically consists of several steps: capturing infant movements, extraction of movement features, and then classification.[14] The extraction of movement data and GMA prediction typically relies on AI models, with pose estimation emerging as the most used method for extracting infant movement data, with several open-source applications available.[25–27] Classification of GMA rather than outcome of CP has been the focus of most research, as classification is available in a timelier manner for the GMA, whilst CP is often formally diagnosed at a later age.[28] Further, the GMA has application for screening beyond CP. Pose estimation is a computer vision technique used to track and detect key body points in an image or video, providing their position over time. For the prediction of GMA classification, various machine learning and deep learning approaches have been used, using infant movement features and demographics as input features.[25] Although automated GMA tools have made great advances, most have been unable to progress beyond demonstrating feasibility due to small datasets.[14,18,24,29] To date, limited data-sharing or pooling has occurred due to logistical, legal and ethical requirements.[29] To ensure accuracy, reliability, and generalisability across multiple populations, further validation of current AI algorithms to automate the GMA are required.

The purpose of this paper is to provide a roadmap that outlines the required steps to automate the GMA (referred to as AI-GMA) spanning research to implementation to achieve our goal of universal screening for CP. Our goal is to ensure all children, including those in LMIC, where there is a higher prevalence of CP, have access to early screening by developing a scalable and accessible solution to accessing the GMA.

Methods

The roadmap was developed through the CP360 initiative, a global effort to advance CP research, care, and advocacy, using a structured four-stage process adapted from a modified model of the Stroke Research Roundtables.[30] Fig. 1 summarises the working group process.

Firstly, in January 2024, the co-chairs (AS and AG) led an international interdisciplinary working group including GMA experts (General Movements Trust Board and Members: AG, AS, CM, PM, AB), automated GMA (TT, PM, SV, EP), persons with lived experience (PWLE) (AC, FF, JC), CEOs of CP organisations and clinician researchers (NB, DM, AR-R). This group defined a summit agenda to bring together experts in the field of automated GMA, along with PWLE, and legal expertise in data sharing and AI. Second, in April 2024, we conducted a scoping review to identify experts in the field who have published within the past 5 years and invited them to attend the summit (AI-GMA summit). Invitations were sent to relevant experts between June and July 2024. The third step involved a 2-day summit in September 2024, based in Pisa, Italy.

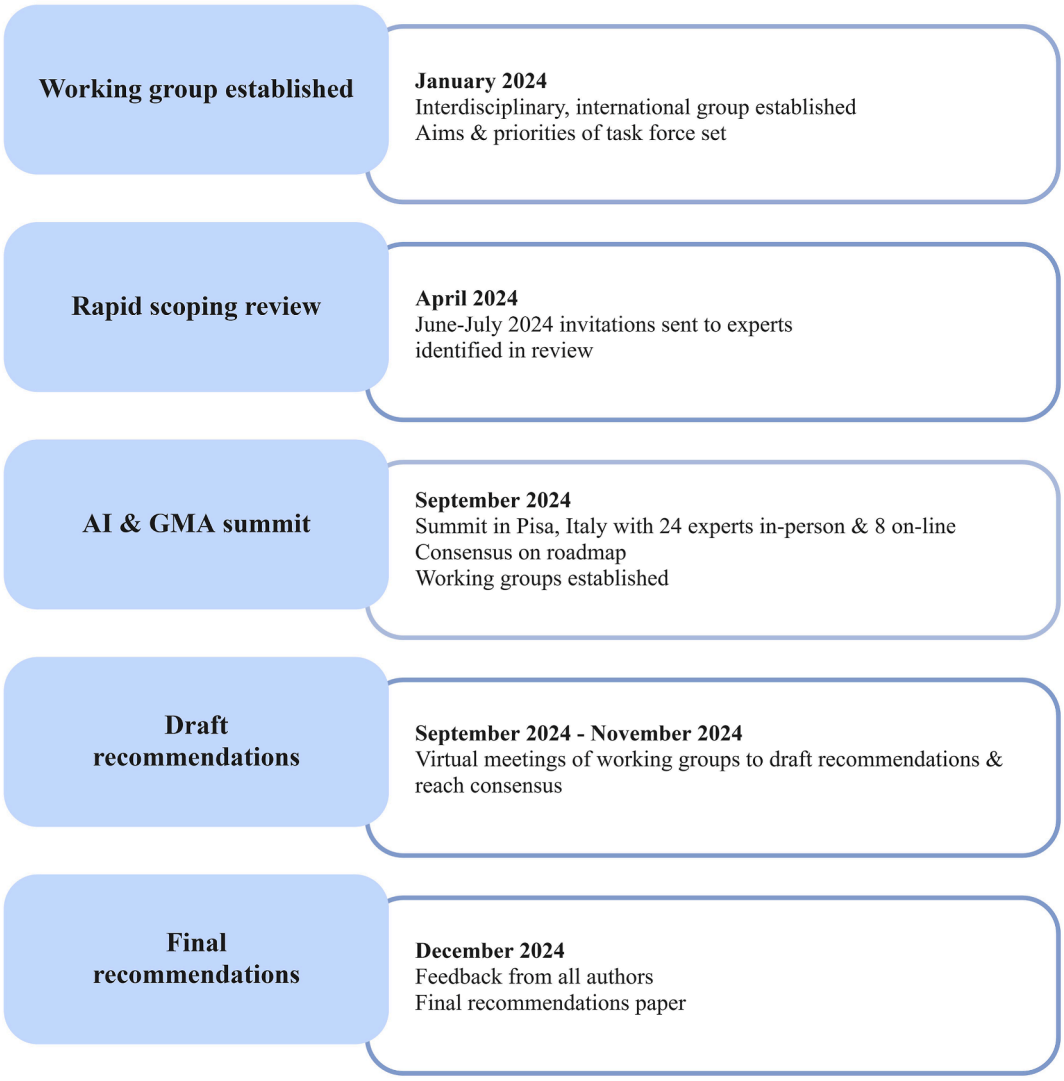

Fig. 1: **Automated GMA-Taskforce process.**

The summit was attended by 32 experts, both in-person and online and the summit was focused on defining the top priorities and developing a roadmap. Working groups were established and face-to-face discussion, with hybrid options for those unable to attend in person, occurred to reach consensus between September and November 2024. For each priority, a subgroup identified opportunities, challenges, and solutions, and reported back to the larger group. When needed, follow-up virtual meetings were conducted to reach consensus. The final stage involved synthesis of the summit and write-up of the roadmap presented herein.

Role of the funding source
The funder(s) of this work had no role in the study design, data collection, data analyses, interpretation, nor the writing of the report.

Results
The AI-GMA summit included wide stakeholder representation from allied health, computer scientists, engineers, lawyers, medical doctors and PWLE. There was global involvement with attendees from Africa, Australia, Europe and North America. Only Asia was not represented by authors who had published in the field. Email invitations sent to corresponding authors went unanswered.

Following discussion of the opportunities, challenges, and key considerations across technical development, clinical integration, ethics and governance, and global implementation, a road map was developed (Fig. 2). Initial steps (Steps 1 and 2) involve a centralised global effort and are detailed in this paper. The final steps are briefly discussed here and will be the focus of future summits (Steps 3, 4 and 5).

Step 1. development of data standards
To enable AI-GMA development, large, diverse datasets of infant movement videos and infant demographics need to be collected with standardised protocols and expert GMA labels.

Data collection and quality
Standardised protocols for video recording are needed to ensure compliance with GMA requirements. Videos can be recorded by caregivers, community workers, or health professionals using handheld mobile devices, provided that suitable instructions are given.[19,21,31,32] Infants should be between 12 and 17 weeks corrected age at the time of filming.[4,33,34] Standard recommendations for GMA recording already exist[8] with Fig. 3 providing a summary of the key video characteristics. Solutions that automatically determine a video's suitability based on filming recommendations and video characteristics should be included in the workflow. Videos that do not meet these requirements may need to be re-recorded. Clinicians' time to review and contact caregivers to request another video can result in the narrow window in which FMs are expressed being missed. Future work should prioritise reducing the time taken to provide feedback on video suitability to caregivers. This could be built into existing video capture applications providing real-time feedback or asynchronously post video upload.

Data labelling
Methods to ensure high-quality, consistent expert labelling of GMA classification and infant demographics are crucial to progress development and ensure interoperability of both data and AI models. GMA classification for the fidgety age period, should include 'Normal FMs', 'Absent FMs', 'Abnormal exaggerated FMs' and 'not scorable' and is the minimum requirement.[33] Alternatively, clip-level annotations that identify the periods of time within a video where FMs are present may offer additional information in the development of AI models.[35,36]

Alongside GMA labels, infant demographic data should be collected, at a minimum including the infant's corrected age (weeks) at the time of filming, gestational age, sex and country of birth.[34,37] If corrected age cannot be obtained, such as in LMIC, then approximate at time of assessment should be included. This minimum data is essential, as these variables have been shown to be related to the GMA classification. Additional information as listed in Table 1 is beneficial but may be more difficult to collect. Both the minimum and additional information have the potential to be included in AI prediction models to improve accuracy for later prediction of CP.

Our roadmap focuses on screening for CP probability using GMA, while other studies have used GMA videos to predict CP diagnosis directly.[10,16,38,39] We have chosen GMA classification as the outcome rather than CP, due to challenges in follow-up.[40] In LMIC where the average age for CP diagnosis is above two years.[28] To ensure effective universal screening in the future, it is important to connect AI-GMA with local CP studies and/or registers, which confirm CP diagnoses from two-five years old. This linkage will help reliably and consistently capture CP outcomes, reduce barriers to long-term follow-up, and establish timeframes that align with local settings.

Step 2. development and validation of AI-GMA from large and diverse data sets
Data consortium
We propose the establishment of a data consortium to provide oversight of dataset curation, model benchmarking and data sharing. Data sharing in this context relates to either direct sharing of videos and demographic data, derived data (i.e. body pose information) or AI Models. The consortium would need to have

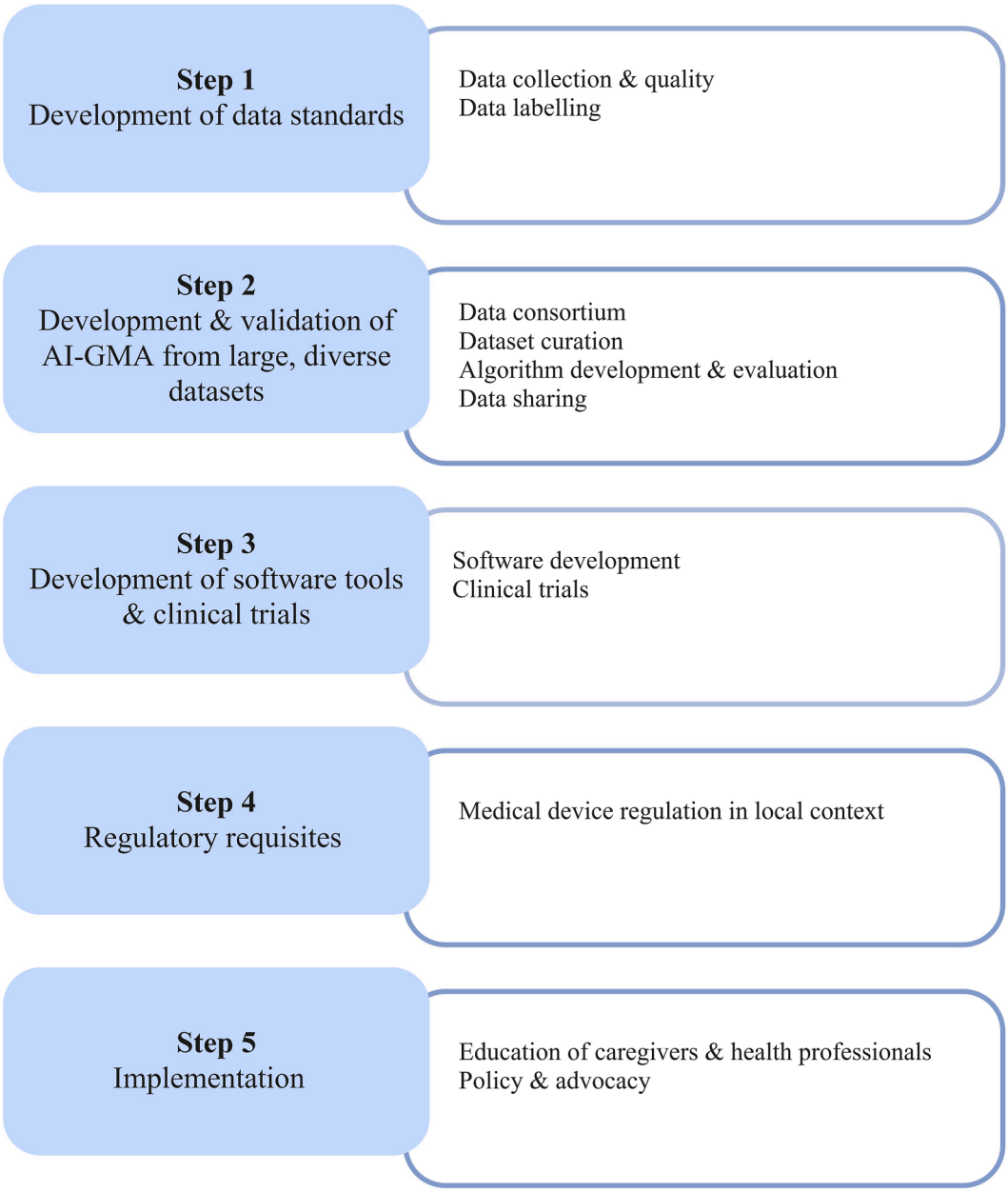

Fig. 2: **Road Map to Universal Screening of CP using AI-GMA**.

a legal agreement between all parties, ensuring the following five principles: (1) The consortium would establish a governance structure responsible for the secure storage of the video datasets; with an executive board which rotates between members. Researchers, clinicians and PWLE will be part of the board, with a global representativeness. (2) Each member would contribute data, expertise or lived experience. Curation of the dataset would ensure quality and diversity of the video data, standardised annotation and scoring, and most importantly compliance with regulations at the individual sites for the collection and storage of the video data. (3) Establishment and maintenance of separate development and benchmarking datasets. (4) Applications for access would be reviewed by the executive board and decided by consensus. A pathway for members to access the dataset would also be established. (5) Each contributor would still maintaining their own data including IP (as relevant) related to their data set.

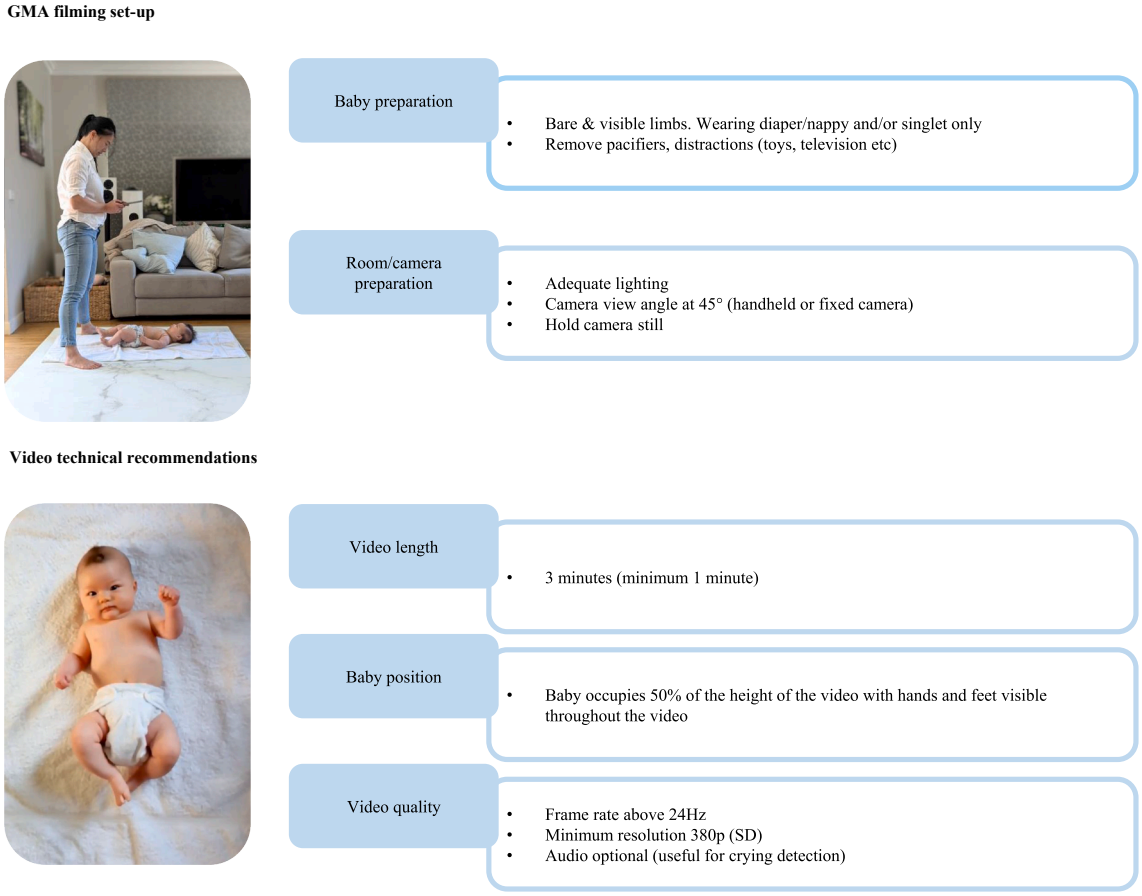

Fig. 3: **Recommendation for recording GMA for AI.**

Dataset curation

Datasets used for the development and benchmarking of AI-GMA must accurately reflect the diversity of the population they are going to be used with in terms of race, ethnicity, context (such as clinical versus home settings), socio-demographics and indicators of CP risk (for example, preterm versus term births). Traditional sample size estimations are generally not applicable to AI model training problems, posing significant challenges in determining the appropriate sample size. Several factors influence the required sample size, including class imbalance, model architecture, and the number of features. A recent study examining sample sizes for AI models in biomedical research showed for comparable studies involving classification problems where the minority class constituted less than 10% of the samples, and using a neural network model, the median sample size required was 13,000.[41] Importantly, the study demonstrated that increasing the representation of the minority class can reduce the required sample size. Existing datasets typically have low representation of absent or abnormal FMs, highlighting the importance of prioritising collection of data from high-risk infants. We recommend the formation of such a dataset through the pooling of data resources from multiple countries and settings. Resources may need to be allocated to support data collection in LMIC to ensure appropriate representation including training of local health professionals

Minimum required dataset
GMA video–Adhering to the recommendation for recording GMA for AI
GMA score–Normal FMs, Absent FMs, Abnormal exaggerated FMs, not scorable
Sex (female, male, unspecified)
Gestational age (weeks)
Age at video recording (term corrected in weeks)
Country of birth
Additional information (optional)
Birth weight (grams)
Birth complications
Multiple or singleton pregnancy
Race
Ethnicity

Table 1: **Dataset variables.**

and systems to store data. A portion of this dataset should be held out to be used for AI model benchmarking. The remaining data can be used to further the development of AI-GMA models. The benchmarking dataset should be randomly selected from the larger dataset while ensuring balanced representation of different ethnic backgrounds, clinical and home settings, and CP risk indicators. The benchmarking dataset should not be used for model development to maintain its integrity as an unbiased test set. These datasets should be ensembled with oversight from the data consortium (see below for further details on the data consortium and data sharing). The data consortium's responsibilities should include curating the data, updating it over time to reflect new population trends or risk indicators, and implementing strict access protocols.

Algorithm development and evaluation. We focus on the development and evaluation of computer algorithms for GMA classification from videos. Both classical machine learning and deep learning approaches have been applied to predict GMA classification. We recommend the establishment and use of a held-out benchmarking dataset to determine the performance of an AI model and its generalisability. For model evaluation, we recommend employing a lock-box method where planned analysis and model features are pre-registered in line with the Open Science Foundation guidelines. Results from the evaluation and the AI model should be reported in a structured fashion using the TRIPOD + AI guidelines and published on a publicly accessible website maintained by the data consortium to ensure transparency.[42]

Performance metrics should include: area under the receiver operator curve, sensitivity, specificity, negative and positive predictive values and balanced accuracy.[43] Balanced accuracy is recommended as the majority of datasets have an unbalanced distribution of GMA labels. In addition, fairness metrics, such as performance across subgroups (e.g. by race and ethnicity, sex or setting), should be established to verify that models are not only accurate but equitable. The model's performance should ideally be consistent across all demographic and probability indicator subgroups. For AI-GMA to be used in clinical settings, they require acceptable psychometric properties. The psychometrics will vary according to local context but are likely to have sensitivity and specificity above 80% when used as a screening tool. Much higher accuracy would been needed for a diagnostic tool, however, AI-GMA is focused on screening and not diagnosis.

Data sharing. AI-GMA currently faces challenges due to a lack of large and diverse datasets, limiting the ability to create robust and generalisable AI models. Collaborative efforts among institutions are essential to address these issues, including data sharing and standardisation of GMA videos.[14,19,29] Ensuring GMA videos meet common data standards will aid in integrating data from different groups. By identifying those videos with standardised protocols and labeling, we can create a more comprehensive and diverse dataset. Proper protocols for secure collection, storage, and sharing of data, especially regarding facial images considered protected health information (PHI), are required for compliance across jurisdictions and should be developed with PWLE. Below we outline three possible solutions for data sharing to progress AI-GMA: a trusted research environment, sharing de-identified data and a federated learning environment.

Trusted research environment. A trusted research environment is a secure digital platform designed to store, manage, and analyse sensitive research data. It allows for controlled access to data, enabling researchers to work with sensitive datasets whilst ensuring privacy, security, and compliance with ethical and regulatory standards.[44] This approach would enable sharing and pooling of GMA videos and infant demographic data. However, it requires significant upfront investment, legal coordination, and ongoing management of infrastructure. It is likely the best option for long-term collaboration, however, large-scale funding is required for it to be sustainable. Additionally, consent from caregivers would be necessary to share videos that include facial images, making this approach more appropriate for prospective data collection efforts.

Sharing of de-identified data. In situations where sharing of PHI is not feasible, such as with existing data where consent has not been obtained for sharing, it may be more appropriate to pool and share de-identified data. Sharing pose-estimation data of the infants' movements captured from GMA videos and limited demographic information could be a viable option. Skeletal or pose data contains either 2D (x,y) or 3D (x,y, and z) information associated with each body point throughout the video and is non-identifiable. Information, such as audio or facial expression is removed and may present challenges in determining if the infant was in the correct behavioral state for GMA. Additionally, the ability to refine pose estimation methods is lost. However, it offers the benefit of sharing beyond a trusted research environment since the data is de-identified. This would enable more groups to contribute to this challenge by making data available through platforms such as Kaggle or PhysioNet.[42] These platforms provide open AI challenges, where groups develop AI models using the provided open access dataset and models are submitted and benchmarked against a held-out dataset.

Blurring of faces in GMA videos has also been explored to de-identify videos. Studies have shown high

agreement on GMA between blurred and non-blurred videos in both human assessors and automated methods.[29] Although some facial information for identifying the infant's behavioral state (e.g. fussiness, drowsiness) may be lost. Screening of these videos for infants' behaviour state prior to deidentification by the local site would be essential to ensure video quality. De-identification by face-blurring retains maximum original data, which enables various further technological explorations that are not dependent on the methods and techniques used to redact data (e.g. extracting poses).

Federated learning. Federated learning provides an alternative to traditional data sharing and is gaining increasing traction in the health setting.[43,44] In federated learning AI models are collaboratively trained across multiple data sources without the need to pool or share data itself. In this approach, each participating site trains a local version of the AI model on its own data, that is then merged to form a global model. Model parameters, and not the data, are shared and merged, which is updated through multiple iterations across sites. This setup preserves data privacy whilst enabling access to broader and more diverse datasets, enhancing the accuracy and generalisability of AI-GMA. However, it requires each site to have sufficient IT infrastructure and computational resources to support local AI model training. Currently due to high graphic processing unit (GPU) costs, federated learning may be inaccessible to some in LMICs but will become more viable as the cost of computational resources for training AI algorithms decreases.

Overall, while face blurring has been explored to some extent, federated learning remains an untested but promising approach for AI-GMA models. Further research is needed to evaluate and compare these methods in terms of sensitivity and diagnostic efficiency.

Step 3. development of software tools and clinical trials

Software development

To date, AI-GMA work has remained in the research domain, relying on batched offline analysis of GMA videos collected in clinical or research settings. To enable universal screening, future efforts should be directed towards development of a software platform that incorporates the full workflow including video upload, storage, real-time analysis and reporting. The platform will also need to include caregiver and clinician facing portals and be scalable. For universal screening to occur it will be important to have AI-GMA incorporated into a product where videos can be obtained by caregivers, health professionals and/or community workers, such as existing mobile or web-based solutions,[19] with the addition of AI-GMA to give timely reporting of results. Rather than the mobile or web-based solution giving feedback directly to families, the results would go to a health professional or appropriate trained personnel to follow a care pathway (Fig. 4). The workflow will need to be co-designed with stakeholders including health professionals and PWLE and adapted to local contexts.[45]

Clinical trials

Extensive clinical trials across diverse demographics of the end-to-end workflow are needed to determine the effectiveness of universal screening for CP. Whilst establishing the effectiveness of AI-GMA models to identify high probability of CP in the real-world is a key indicator, metrics on useability from caregiver and clinician perspective, clinical validity, reproducibility and safety also need to be evaluated. Ideally trials should be designed to gather evidence required to support regulatory approvals for medical devices by the relevant body in each region (e.g. Food and Drug Administration [FDA], Therapeutics Good Act [TGA], European Conformite Europeenne [CE] mark) and need to occur with step 4 below.

Infants from diverse ethnic, race[37] and socioeconomic backgrounds should be included to account for variations in developmental norms and general movement patterns.[46] Conducting trials in multiple geographic locations helps to evaluate model robustness in detecting movement patterns across different environments and healthcare practices.

A phased approach for developing clinical trials should be adopted. Within each phase AI model performance should be benchmarked against experienced GMA assessors who are blinded to the model's predictions to ensure an unbiased comparison. Phase 1 should focus on feasibility, testing basic functionality, and ensuring the AI models can handle variability in video data. Phase 2 should focus on validation in a selection of diverse cohort. The focus here is on validating the model's accuracy and consistency across a broader set of demographics, settings, and CP probability indicators. Phase 3 should include large multisite clinical trials across different regions and populations. The primary aim is to confirm that the AI model and workflow can reliably predict GMA outcomes in a broad, representative sample of infants. Phase 3 trials should use blinded, randomised assessments to mitigate biases and validate the model's generalisability and clinical utility.

Step 4. regulatory requisites

Medical device regulation in local context

It will be necessary to ensure that the product complies with regulatory requirements in the country of use including approvals from regulatory bodies. Regulatory compliance is essential to ensure safety, efficacy, and trust in clinical adoption, noting that regulation practices vary around the world and are continuously being updated. AI-GMA would support clinical decision-

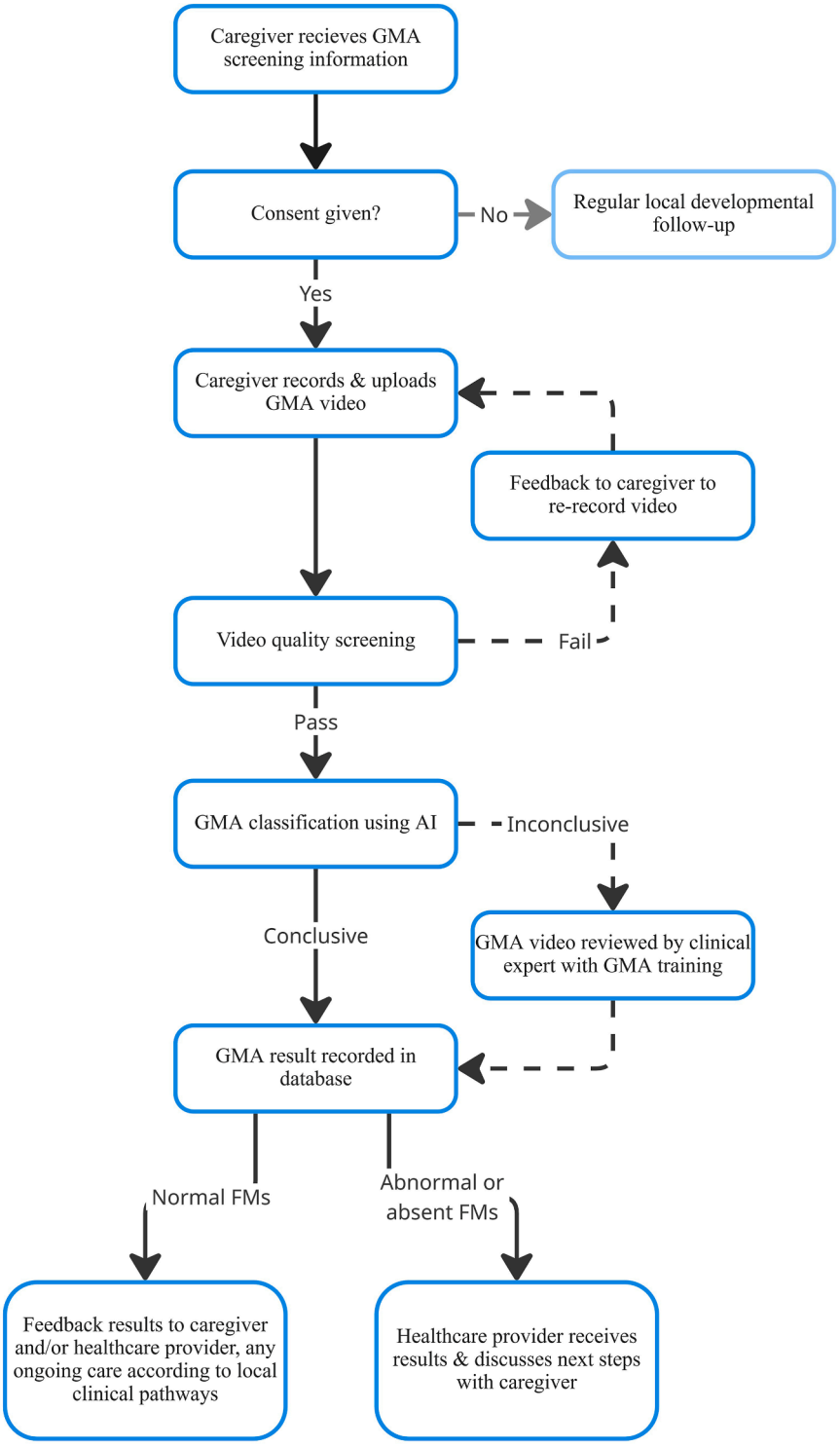

Fig. 4: **Possible workflow for AI-GMA.**

making, likely classifying it as a medical device. This classification requires stringent safety and clinical testing, with ongoing verification and validation to ensure accuracy. Compliance with regulatory acts, such as the AI Act Europe requires transparency, human oversight, and security measures and is essential to

establishing its credibility as a decision-making tool. Adherence to the General Data Protection Regulation (GDPR) in Europe, Health Insurance Portability and Accountability Act (HIPAA) in the US, and other local regulations is crucial and made more complex by the use of data with identifying features such as facial features. Regulatory standards require post-market monitoring, essential for AI devices that may need continuous updates. Collecting feedback, managing adverse event reports, actively engaging primary stakeholders and updating algorithms are vital to maintaining compliance and user trust. Regulations also emphasise ethics and transparency, requiring AI decisions to be interpretable, justifiable, and unbiased.

Step 5. implementation

Education of caregivers and health professionals

The implementation pathway will need to be adapted to the local setting, with training of health professionals and caregivers to understand why early screening is important and the pathway for families from consent to diagnosis. Feedback will need to be given in a timely manner with an understanding of sensitivity and specificity of screening using AI-GMA. Standardised pathways for further assessment and early intervention must be established for infants flagged as high probability of CP. Recommended diagnostic pathways exist for LMICs, but implementation of these pathways varies by setting.[3] The proposed tool will be a screening or decision support tool rather than a diagnostic tool and this important distinction should be underscored when the global community is educated on its use.

Policy and advocacy

Engagement with policymakers and health systems to integrate AI-GMA into standard care is needed at all stages of the roadmap but particularly for implementation. It is important to develop guidelines for universal screening by working with global health organisations for best practices in implementing and using the automated AI-GMA tool. Public awareness campaigns can help educate the public and healthcare providers about the importance of early CP detection and the availability of the AI-GMA screening tool. Further, it will be essential to work with local governments and patient and disability associations to make universal screening for CP a part of standard neonatal care protocols, ensuring it is covered by public health policies and insurance plans under the principle of universal health coverage.

Discussion

There has been a rapid increase in the number of publications on automating the GMA for early detection of CP in the past decade.[18] To achieve our goal of AI-GMA with a focus on health equity and access, large data sets of GMA recordings are needed, and families need to be assured that data privacy and regulatory requirements are met. This involves trust and co-design in the user experience and long-term protection of their child's data. The development and implementation of AI-GMA must be informed by perspectives and experiences of individuals with CP and their families. Our summit invited experts from around the globe, including LMIC, to reach consensus on a road map to progress the field and included parents of children with CP who emphasised the importance of data ownership, co-design, consent, trust and communication. Although AI has the potential to increase accessibility to early screening, health professionals will always need to be involved in the pathway to diagnosis; from consenting families for early screening to final diagnosis of CP which involves several assessments (e.g. neurological examination, clinical history, neuroimaging).[4,12,45] Further, the use of AI should not supersede the need for close follow up for infants and where there are clear indicators for a CP diagnosis already such as major brain injury.

GMA alone are not a diagnostic tool for CP, rather a disruption to the central nervous system requiring further investigation. Advancements in GMA automation has potential to facilitate early detection of not just CP but also other neurodevelopmental delays.[20,47] Small studies have shown that motor repertoires observable within a GMA video can predict increased probability of cognitive disability.[20,48,49] This points to the possibility for the GMA to predict outcomes beyond CP and identify those children who might also benefit from early intervention for other conditions. Inspired by the development of AI-GMA, the authors see the possibility for more research to enhance AI algorithms that automate early detection of conditions like autism spectrum disorder and attention deficit hyperactivity disorder based on movement data from the early motor repertoires of infants.

The use of AI-GMA is not without its limitations. Trust in AI, data privacy and security and the importance of human connection when making a diagnosis are just some of the many considerations in making universal screening possible.[50] With this roadmap, we aim to progress both research and clinical application of AI and GMA whilst acknowledging the rapid changes in the field. Substantial capital investment will be needed to achieve each step in the roadmap and it is essential to collaborate to reduce duplication of efforts and targeting resources appropriately. Ongoing summits which will expand on these issues will be required.

Conclusions

AI-enabled GMA has immense potential to allow for universal early detection of CP, but realising this potential requires a coordinated, multidisciplinary effort.

Search strategy and selection criteria

References for this Health Policy were identified through searches of PubMed and Google Scholar with the search terms "general movements" AND "artificial intelligence" OR "machine learning" OR "computer vision" OR "automated" OR "deep learning" AND "cerebral palsy" from 2017 until February 2024. Articles were also identified through reference lists of key review papers. No language restrictions were used. The final reference list was generated on the basis of originality and relevance to the broad scope of this Health Policy.

This review represents a consensus roadmap from a diverse group of stakeholders. We call for the formation of a global effort to advance AI-GMA development and implementation, with a focus on equity and improving outcomes for all children with high probability of CP worldwide. By addressing these key areas, we can make automated universal screening for CP using the GMA available to everyone, leading to earlier intervention and better outcomes for children worldwide.

Contributors

AJS, PBM, NB, RB, JC, AG, AMcE, CM, DM, AR-R, TT, and EP were responsible for conceptualisation. AJS and EP handled data curation and the formal analysis. AJS, NB, RB, and AG acquired funding. AJS PBM, LA, NB, RB, AFB, AC, JC, FF, AG, ESLH, MJJ, AMcE, CM, AM, DMM, SO, CP, LAP, AR-R, TT, DZ, and EP were involved in investigation. AJS, PBM, NB, RB, AG, JC, AMcE, DMM, AR-R, TT, EP participated in study methodology. AJS, EP, AK were responsible for project administration. AJS, AK, LAP, EP handled visualisation. AJS, PBM, LA, ESLH, AE, CM, DMM, SO, CP, LAP, DZ, EP wrote the original draft of the manuscript. AJS, PB, LA, NB, RB, AFB, AC, JC, FF, AG, ESLH, MJJ, AK, AMcE, CM, AM, DMM, SO, CP, LAP, AR-R, TT, DZ, and EP were involved in subsequent review and editing of the manuscript.

Declaration of interests

AJS, PBM, AB, AG, CM, and CP are all members of the General Movements Trust. Norwegian University of Science and Technology (NTNU) and St. Olavs Hospital, Trondheim University Hospital, may benefit financially from a commercialization of the In-Motion AI-tools through existing intellectual properties which may include LA. All other authors declare no competing interests.

Acknowledgements

This meeting was supported by grants from Cerebral Palsy Foundation, Cerebral Palsy Alliance and European Union Born to Get There for coordination of the summit, accommodation and catering and flights for participants from low-middle income/low resource settings. AJS was funded by the National Health and Medical Research Council (NHMRC) Investigator Grant 2033694. PBM and DZ were supported by the VW-IDENTIFIED and DFG-456967546 projects. AG was supported by the Italian Ministry of Health–Ricerca Corrente (2025). LA was supported by grants from The Liaison Committee between the Central Norway Regional Health Authority and the Norwegian University of Science and Technology, Trondheim, Norway (SO-ID: 24239). EH is supported by the Engineering and Physical Sciences Research Council (EPSRC) and Medical Research Council (MRC) Impact Acceleration Account Standard Project. EP is supported by Murdoch Children's Research Institute Clinical Science Fellowship.

References

1 Bax M, Goldstein M, Rosenbaum P, et al. Proposed definition and classification of cerebral palsy, April 2005. *Dev Med Child Neurol*. 2005;47(8):571–576.
2 McIntyre S, Goldsmith S, Webb A, et al. Global prevalence of cerebral palsy: a systematic analysis. *Dev Med Child Neurol*. 2022;64(12):1494–1506.
3 Morgan C, Fetters L, Adde L, et al. Early intervention for children aged 0 to 2 years with or at high risk of cerebral palsy: international clinical practice guideline based on systematic reviews. *JAMA Pediatr*. 2021;175(8):846–858.
4 Novak I, Morgan C, Adde L, et al. Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment. *JAMA Pediatr*. 2017;171(9):897–907.
5 te Velde A, Morgan C, Novak I, Tantsis E, Badawi N. Early diagnosis and classification of cerebral palsy: an historical perspective and barriers to an early diagnosis. *J Clin Med*. 2019;8(10).
6 Maitre NL, Damiano D, Byrne R. Implementation of early detection and intervention for cerebral palsy in high-risk infant follow-up programs: U.S. and global considerations. *Clin Perinatol*. 2023;50(1):269–279.
7 King AR, Al Imam MH, McIntyre S, et al. Early diagnosis of cerebral palsy in low- and middle-income countries. *Brain Sci*. 2022;12(5).
8 Einspieler C, Prechtl HF. Prechtl's assessment of general movements: a diagnostic tool for the functional assessment of the young nervous system. *Ment Retard Dev Disabil Res Rev*. 2005;11(1):61–67.
9 Prechtl HF, Einspieler C, Cioni G, Bos AF, Ferrari F, Sontheimer D. An early marker for neurological deficits after perinatal brain lesions. *Lancet*. 1997;349(9062):1361–1363.
10 Kwong AKL, Fitzgerald TL, Doyle LW, Cheong JLY, Spittle AJ. Predictive validity of spontaneous early infant movement for later cerebral palsy: a systematic review. *Dev Med Child Neurol*. 2018;60(5):480–489.
11 Spittle AJ, Doyle LW, Boyd RN. A systematic review of the clinimetric properties of neuromotor assessments for preterm infants during the first year of life. *Dev Med Child Neurol*. 2008;50(4):254–266.
12 Kwong AKL, Eeles AL, Anderson PJ, et al. The knowledge translation of early cerebral palsy (KiTE CP) study: implementing screening among a high-risk prospective cohort of Australian infants. *J Pediatr*. 2024;268:113949.
13 Reich S, Zhang D, Kulvicius T, et al. Novel AI driven approach to classify infant motor functions. *Sci Rep*. 2021;11(1):9888.
14 Silva N, Zhang D, Kulvicius T, et al. The future of General Movement Assessment: the role of computer vision and machine learning–a scoping review. *Res Dev Disabil*. 2021;110:103854.
15 Passmore E, Kwong AL, Greenstein S, et al. Automated identification of abnormal infant movements from smart phone videos. *PLOS Digit Health*. 2024;3(2):e0000432.
16 Groos D, Adde L, Stoen R, Ramampiaro H, Ihlen EAF. Towards human-level performance on automatic pose estimation of infant spontaneous movements. *Comput Med Imaging Graph*. 2022;95:102012.
17 Gao Q, Yao S, Tian Y, et al. Automating General Movements Assessment with quantitative deep learning to facilitate early screening of cerebral palsy. *Nat Commun*. 2023;14(1):8294.
18 Deng W, O'Brien MK, Andersen RA, Rai R, Jones E, Jayaraman A. A systematic review of portable technologies for the early assessment of motor development in infants. *NPJ Digit Med*. 2025;8(1):63.
19 Marschik PB, Kwong AKL, Silva N, et al. Mobile solutions for clinical surveillance and evaluation in infancy-general movement apps. *J Clin Med*. 2023;12(10).
20 Kwong AKL, Doyle LW, Olsen JE, et al. Parent-recorded videos of infant spontaneous movement: comparisons at 3-4 months and relationships with 2-year developmental outcomes in extremely preterm, extremely low birthweight and term-born infants. *Paediatr Perinat Epidemiol*. 2022;36(5):673–682.
21 Adde L, Aberg KB, Fjortoft T, et al. Implementation of remote general movement assessment using the in-motion instructions in a high-risk Norwegian cohort. *BMC Pediatr*. 2024;24(1):442.
22 Peyton C, Frazier M, Aaby D, et al. Assessing neonatal intensive care unit (NICU) graduates across varied settings: a study on the feasibility of the baby moves app. *J Pediatr*. 2024;269:113979.
23 Spittle AJ, Olsen J, Kwong A, et al. The Baby Moves prospective cohort study protocol: using a smartphone application with the General Movements Assessment to predict neurodevelopmental

outcomes at age 2 years for extremely preterm or extremely low birthweight infants. *BMJ Open*. 2016;6(10):e013446.
24 Irshad MT, Nisar MA, Gouverneur P, Rapp M, Grzegorzek M. AI approaches towards prechtl's assessment of general movements: a systematic literature review. *Sensors (Basel)*. 2020;20(18).
25 Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y. OpenPose: realtime multi-person 2D pose estimation using Part Affinity fields. *IEEE Trans Pattern Anal Mach Intell*. 2021;43(1):172–186.
26 Mathis A, Mamidanna P, Cury KM, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci*. 2018;21(9):1281–1289.
27 Jahn L, Flugge S, Zhang D, et al. Comparison of marker-less 2D image-based methods for infant pose estimation. *Sci Rep*. 2025;15 (1):12148.
28 Khandaker G, Muhit M, Karim T, et al. Epidemiology of cerebral palsy in Bangladesh: a population-based surveillance study. *Dev Med Child Neurol*. 2019;61(5):601–609.
29 Marschik PB, Kulvicius T, Flugge S, et al. Open video data sharing in developmental science and clinical practice. *iScience*. 2023;26(4):106348.
30 English C, Simpson DB, Billinger SA, et al. A roadmap for research in post-stroke fatigue: consensus-based core recommendations from the third Stroke Recovery and Rehabilitation Roundtable. *Int J Stroke*. 2024;19(2):133–144.
31 Kwong AK, Eeles AL, Olsen JE, Cheong JL, Doyle LW, Spittle AJ. The Baby Moves smartphone app for General Movements Assessment: engagement amongst extremely preterm and term-born infants in a state-wide geographical study. *J Paediatr Child Health*. 2019;55(5):548–554.
32 Kwong AKL, Eeles AL, Olsen JE, Zannino D, Kariotis T, Spittle AJ. Instructional guides for filming infant movements at home are effective for the General Movements Assessment. *J Paediatr Child Health*. 2022;58(5):796–801.
33 Einspieler C, Peharz R, Marschik PB. Fidgety movements–tiny in appearance, but huge in impact. *J Pediatr*. 2016;92(3 Suppl 1):S64–S70.
34 Kwong AKL, Olsen JE, Eeles AL, et al. Occurrence of and temporal trends in fidgety general movements in infants born extremely preterm/extremely low birthweight and term-born controls. *Early Hum Dev*. 2019;135:11–15.
35 Morais R, Le V, Morgan C, et al. Robust and interpretable general movement assessment using fidgety movement detection. *IEEE J Biomed Health Inform*. 2023;27(10):5042–5053.
36 Morais R, Tran T, Alexander C, et al. Fine-grained fidgety movement classification using active learning. *IEEE J Biomed Health Inform*. 2024;29(1):596–607.
37 Einspieler C, Bos AF, Spittle AJ, et al. The general movement optimality score-revised (GMOS-R) with socioeconomically stratified percentile ranks. *J Clin Med*. 2024;13(8).
38 Bosanquet M, Copeland L, Ware R, Boyd R. A systematic review of tests to predict cerebral palsy in young children. *Dev Med Child Neurol*. 2013;55(5):418–426.
39 Groos D, Adde L, Aubert S, et al. Development and validation of a deep learning method to predict cerebral palsy from spontaneous movements in infants at high risk. *JAMA Netw Open*. 2022;5(7): e2221325.
40 J O, McGinley JL, Fox LM, Spittle AJ. Challenges of neurodevelopmental follow-up for extremely preterm infants at two years. *Early Hum Dev*. 2015;91(12):689–694.
41 Silvey S, Liu J. Sample size requirements for popular classification algorithms in tabular clinical data: empirical study. *J Med Internet Res*. 2024;26:e60231.
42 Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000;101(23):E215–E220.
43 Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med*. 2020;3:119.
44 Kavianpour S, Sutherland J, Mansouri-Benssassi E, Coull N, Jefferson E. Next-generation capabilities in trusted research environments: interview study. *J Med Internet Res*. 2022;24(9):e33720.
45 Cameron KLI, Coulston F, Kwong AK, et al. Parents' experiences of early screening for cerebral palsy: a qualitative reflexive thematic analysis. *Dev Med Child Neurol*. 2025;67(6):788–801.
46 O'Connor MI. Equity 360: gender, race, and ethnicity-the power of AI to improve or worsen health disparities. *Clin Orthop Relat Res*. 2024;482(4):591–594.
47 Elliott C, Alexander C, Salt A, et al. Early Moves: a protocol for a population-based prospective cohort study to establish general movements as an early biomarker of cognitive impairment in infants. *BMJ Open*. 2021;11(4):e041695.
48 Einspieler C, Bos AF, Libertus ME, Marschik PB. The general movement assessment helps us to identify preterm infants at risk for cognitive dysfunction. *Front Psychol*. 2016;7:406.
49 Olsen JE, Cheong JLY, Eeles AL, et al. Early general movements are associated with developmental outcomes at 4.5-5 years. *Early Hum Dev*. 2020;148:105115.
50 Ning Y, Teixayavong S, Shang Y, et al. Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. *Lancet Digit Health*. 2024;6(11):e848–e856.