Review Article

General movement assessment by machine learning: why is it so difficult?

William Schmidt[1], Matthew Regan[2], Micheal Fahey[2], Andrew Paplinski[1]

[1]Faculty of IT, Clayton Monash University, Melbourne, Australia; [2]Monash Health, Melbourne, Australia

Contributions: (I) Conception and design: All authors; (II) Administrative support: All authors; (III) Provision of study material or patients: None; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: Schmidt; (VII) Final approval of manuscript: All authors.

Correspondence to: William Schmidt. Faculty of IT, Clayton Monash University, Melbourne, Australia. Email: william.schmidt@monash.edu.

Abstract: The current rate of cerebral palsy (CP) per live births in Australia is between 0.14% and 0.2%, worldwide the rate has been static for 60 years at 0.2%. Typically a CP diagnosis is delayed until around age 2 years; this delay decreases the likelihood of a long-term positive patient outcome. Current early detection is by visual examination of newborns 10 to 20 weeks post gestation. A screening program based on filming babies and processing the video via artificial intelligence (AI) will allow increased early detection and intervention. This paper outlines the practical development, and initial results from, a recurrent deep neural net solution for the classification of newborn videos, specifically targeting CP, using the largest fidgety movements dataset in Australia.

Keywords: Cerebral palsy (CP); fidgety movements; early intervention; deep neural networks; convolution neural networks; long short-term networks

Received: 13 May 2019; Accepted: 27 May 2019; published: 18 July 2019.

doi: 10.21037/jmai.2019.06.02

View this article at: http://dx.doi.org/10.21037/jmai.2019.06.02

Introduction

The diagnosis's of neurological conditions is a difficult and confounding task carried out by specialist practitioners. The aim of this work is not to displace the specialist but to provide an early detection filtering tool to assist in the better allocation of specialist resources. As neurological conditions are confounded by a significant number of factors, the conditions aetiology, genetics, trauma, nutrition, socio-economic conditions and resources (1), their diagnosis is a resource intense process completed over a period of time.

The advent of artificial intelligence (AI) algorithms and wide spread relatively low cost computing power, has provided the opportunity to transfer specialist knowledge to machine based platforms allowing the use of this knowledge by a wider audience. Making simple screening tools readily available allows better resource allocation. The intention of this project is to develop and demonstrate a methodology to produce such low cost diagnostic tools. The paper is constructed as follows: Introduction to cerebral palsy and its diagnosis; literature review, examining AI algorithms that have been proposed and tested for similar diagnostic testing tools; data sources used for in this research; discussion of initial DNN model architecture; training and testing results; discussion of results and; proposed further research program.

Cerebral palsy (CP)

CP is an umbrella term for a range of cerebral disorders all attributed to disturbances of the developing foetal or infant brain. CP mostly originates from a brain injury event occurring before the age of 6 months corrected age. A CP diagnosis requires motor dysfunction, a cerebral injury, no progression of the injury and involvement of the central nervous system with other neurologic and behavioural disorders (2). Genetic pre-dispositions, maternal disease, premature birth, low birth weight and asphyxia during birth are all indicators of increased risk of CP (2-4). Australia is one of only a few countries that have a CP specific register

© Journal of Medical Artificial Intelligence. All rights reserved.

for the monitoring of the condition in the community. The current rate of CP per live births in Australia is between 0.14% and 0.2%, worldwide the rate has been static for 60 years at 0.2%, the rate of CP increases with increase in pre-term birth rising to 14.6% in children born at 20 to 27 weeks gestation (4,5). The annual cost of per person with CP is estimated to be between $43k and $115k ranging with severity of disability (6).

Early diagnoses of CP is considered essential because it enables specific interventions at times when neural plasticity maximises possible gains (7,8). However to date there is minimal evidence supporting the efficacy of early intervention, Hadders-Algra, Herskind and Granild-Jensen (3,4,9,10) argue that this is a consequence of studies on early intervention concentrating on children at high risk of CP and is not indicative of more general cases and that further study is required. Blauw-Hospers, Hadders-Algra (11) reviewed 34 studies of early intervention and found that specific motor training programmes, such as training of locomotor movements on a treadmill or interventions aimed at stimulation of active motor response have been shown to have a positive effect on development.

Diagnosis of CP

The typical course of a CP diagnosis begins with a risk assessment where, pre-term children, those born with birth weight <1,500 g are considered high risk. Early symptoms include abnormal muscle tone, epilepsy, gastro-oesophageal reflux and feeding difficulties. There are many casual mechanisms for CP including, perinatal stroke, hypoxic ischemic encephalopathy (HIE), infection, and birth asphyxia (12). Hadders-Algra (3) notes that the most common lesion in children with CP is damage to the periventricular white matter (19% to 45% of cases), and grey matter injuries including the basal ganglia, thalamus (21%) and cortical infarcts (10%). Scanning techniques such as cranial ultrasound and brain magnetic resonance imaging (MRI) are strong predictors of CP with sensitivity in the range 86% to 100% and specificity in the range of 89% to 97% in studies of high risk children (13). Aisen *et al.* (14) note that for up to 17% of children with CP imaging fails to detect any abnormalities.

Most children born with CP are not born premature and will go undiagnosed until abnormal development is observed. Typically this occurs between 13 months, on detection by paediatricians or 27 months, instigated by parental concern (15). For low risk children detection is generally through developmental observation. Researchers have identified the Fidgety movement period of general movement (GM) development with the course of CP (16). GMs consist of gross motor movements of variable speed and amplitude involving all parts of the body but lacking any sequence of the body parts. GMs appears in the foetus at around 28 weeks postmenstrual age (PMA) in the womb before individual limb movements are observed. From 38 weeks GMs are observed to change to a writhing pattern which continues until around 8 weeks post term when the writhing movements are replaced by fidgety movements.

Fidgety movements (FMs)

FMs are characterised by small velocities and amplitudes with movement of neck, trunk and limbs in all directions, and are seen continually in awake children who are not fussing or crying (13). FMs may be seen as early as 6 weeks but typically occur around 9 weeks and fade out at around 20 weeks post term. FMs are most reliably detected at 12 to 14 weeks post term age (17).

Fidgety movements may occur in isolation or increase and decrease in frequency as they traverse the body. FMs generally occur in all body parts, although not simultaneously. The FMs often occur with other movements such as wiggling-oscillations and saccadic arm movements, leg lifting, hand-knee contact, trunk rotation and axial rolling. The challenge of detecting FMs is thereby increased by the presence of these gross motor movements (17).

General movements assessment (GMA)

Current studies indicate that the GMA is the most sensitive and specific test available to allow early detection of CP. Studies show GMA has sensitivity and specificity of 98% and 91%, compared with cranial ultrasound 74% and 92%, neurological examination 88% and 87%, while MRI ranges from 86% to 100% and 89% to 97% respectively (4). GMA does not provide a prediction to severity of the condition (13,17). The GMA is applied through gestalt observation of movements by a trained observer, the process involves looking for abnormal GMs, characterised by limited variation and limited variability in GM, the presence of cramped synchronised general movements (CSGMs) and the absence of fidgety movements are strong indicators of abnormal GMs. CSGMs lack fluency and complexity they appear stereotyped, with limb and trunk muscles contracting and relaxing almost simultaneously.

© Journal of Medical Artificial Intelligence. All rights reserved.

GM assessment requires considerable practise and training before an assessor reaches a proficiency level that provides reliable and accurate evaluations. The GMA requires the child to be in an alert awake state, but not stressed or crying. Most accurate assessments, for the prediction of CP, occur with the child between 12 and 20 weeks, which is less than optimum from an early intervention and neural plasticity perspective (3,18).

Current work

The proposal of this project is to transfer expert judgment, in terms of ability to recognise and rate the GMs, specifically FMs, in neo-nates to an AI based diagnostic tool. Clinicians require significant levels of training and experience to obtain levels of testing reliability suitable for use in clinical settings. Current machine learning techniques, in particular the field of deep learning, do not provide a panacea that replaces experience and judgment; however, they can supplement and augment those judgements. Automation of diagnostic testing can improve levels of test reliability through reduction in variation and tester biases, however automation it is also associated with a perceived loss of autonomy and clinical judgment (19). It is the intention of this project to provide these sorts of benefits while being aware of the limitations of the proposed technologies.

Directed feature selection

One of the premises of this project is that the current diagnostic techniques of the GMA examination can be mapped to machine learning (ML) based classifiers, as seen in the papers reviewed below. There are two basic approached to the mapping problem, either allow the ML algorithm to identify relevant features and map these to the diagnostic output, or pre-select features and use these to map to the ML classifier. Rahmati *et al.* (20) argue that when there are relatively few subjects but multiple possible features, allowing the ML classifier to select features frequently leads to a suboptimal solution. The power frequency spectrum of 78 infants' post-term age 10 to 18 weeks was examined using video capture and six accelerometers attached to the limbs of the children, to determine a suitable set of features, and to compare data sources. An optical flow algorithm, as given in (21), was used to calculate flow vectors for body segments. A fast Fourier transform, FFT, analysis was done of both the video flow vectors and the accelerometer data, and a cross-validation algorithm was applied to select significant features, frequency bands, from the power frequency spectrum. The frequency bands between 25 and 35 Hz were found to be most significant, Rahmati *et al.* (20) interpret this as the 25 Hz being associated with slow translational movement while the 35 Hz is associated with abrupt changes in motions, jerking.

Stahl *et al.* (22) also used optical flow to derive body part flow vectors from video of infants (15 with CP and 67 without). Wavelet analysis, an extension of FFT's, was used to obtain a wavelet power spectrum from the flow vectors. From this data three sets of features were derived absolute motion distance, relative frequency and magnitude of wavelet coefficients. These were then presented to a support vector machine (SVM) for classification. Features were presented individually and in combinations to the SVM to determine the most discriminatory set of features. It was found that absolute motion distance and relative frequency provided better discrimination than wavelet coefficients.

Friedman *et al.* (23) noted that infants at risk of developing CP often also suffer from vocalisation issues and prosed using vocalisation recordings in combination with video and depth data to derive classification features. A number of researches (18,23-26) looking at the CP classification problem have used the Microsoft Kinect system, which provides both an RGB video and a distance to camera, depth, channel. The system was developed for Xbox play interactions and is quoted as a simple hardware system for depth data gathering; it is designed around children with bodies greater than 1m tall and appears to be less effective with smaller bodies. Friedman *et al.* derived a set of descriptive statistics from the data and used these as input features to a step wise linear discriminant analysis algorithm for subject classification. Forty-one infants were examined, with 18 classified as normal neurological development and 23 as abnormal, the classifier was found to have a performance accuracy of 87% (sensitivity 73%, specificity 98%) compared to clinical experts. The study results also verified the correlation between motor and vocal features during neuro development.

Ansari *et al.* (27) present a system for capturing infant motion to monitor and classify Hammersmith Infant Neurological Examinations (HINE). This system involves videoing the infants while carrying out exercises and then identifying the exercises performed. Four exercises are being classified, pulled to sit, vertical suspension, lying horizontal and lying vertical. The video is pre-processed by applying the SIFT algorithm to each frame. The features detected by the SIFT algorithm, across multiple frames, are

© Journal of Medical Artificial Intelligence. All rights reserved.

then clustered using a K-means algorithm and a descriptor, a visual word for the K-means clustered SIFT points, was generated. Features for classification were then generated by treating the descriptors as so many words in a collection of documents and applying a bag of words (BOW) algorithm to generate a descriptor for the frame sequence. These sequence descriptors were than used as inputs to train a Hidden Markov Model, HMM, to classify the sequences according to the four exercises. It was found that this approach could not distinguish pulled to sit, and lay vertically well, so a second step was added to improve the classification. A final classification accuracy of 84% was achieved.

Interestingly, the additional step added by Ansari *et al.* to improve classification involved using skin colour to segment the infant's movement from the image. The frames were converted from the RGB colour space to the YCrCb color space as the Y, luminance component of the colour space was found not to affect the segmentation process making skins of different complexions easier to detect. In total Ansari *et al.* (27) applied seven different computer vision (CV) algorithms to process the video data prior to classification using the HMM approach.

Khan *et al.* (28) developed a tool for the monitoring of Vojta therapy in the home. Votja therapy, applied with infants suspected of suffering CP, is a reflex nerve simulation technique intended to stimulate natural reflexes that appear suppressed in CP children (28). The intention was to use this as a feedback tool to improve the quality of therapy provided by the child's parents. As with the earlier papers the approach uses both video and depth data of infants undergoing therapeutic treatment, and seeks to isolate handpicked features from the data stream for presentation to a classifier, a SVM in this example. The performance of three algorithms for pre-processing and segmenting the child's image into major body components are compared. Bounding boxes are used to identify the total movement of segmented areas of the images such that the area of the box is proportional to the level of movement observed. Nine geometric ratio features are derived to describe the bounding boxes and these are presented for classification to the SVM. A 5-fold cross-validation was performed to validate the system, with 10 subjects, and found to be classifying at around the 80% level (29).

The work reviewed so far has concentrated on two main aspects of the problem, analysing the video to extract motion data and classifying the extracted data. To retrieve motion data a number of algorithms have bene used and tested, all with varying degrees of success. Classification has been based on using a subset of features selected from the images by application of a particular CV algorithm. The chosen features are then presenting to ML classifiers, such as SVM and HMM. An alternative approach which has also been applied is to allow the classifying algorithm to decide what constitutes the features of interest of the image and present these to the classifier.

Un-directed feature selection

Li *et al.* (30) trained a three layer convolutional neural network (CNN) of (C6,13),p,(C72,9),p,fc configuration, using the back propagation stochastic gradient decent (SGD) algorithm. The CNN generated a score for all possible object locations in a frame with the highest score location being chosen as the object's location. In order to use the CNN for a tracking task the SGD algorithm was modified by adding a temporal element to the training data. The temporal element is achieved by drawing the mini batch samples from the positive and negative results sets assuming a different distribution for each set. Testing of this algorithm achieved a tracking success rate of 83% accuracy while the state-of-the-art, in 2014, was 74%, compared to the skin colour modelling tracking method.

Wang *et al.* (31) investigated the use of CNN's in general object tracking. A VGG-16 CNN trained on ImageNet database was examined to identify the relationship between the data presented at the receptor field (inputs) and the activation fields of each layer of the CNN. Wang *et al.* [2015] observed (I) activated feature maps tend to be sparse and localised compared to the receptor field. (II) The feature maps are noisy and do not discriminate a target from its background. (III) Lower layers appear to encode features for discrimination between object classes while the higher layers encode the overall concept of the object and its background.

Wang *et al.* (31) used these observations to develop their tracking algorithm. The proposed algorithm takes the outputs of the two layers, 4-3 and 5-3 and passes these through, two two-layer CNN's, labelled SNet and GNet. The SNet layers occur early in the CNN and learn filters associated with basic edge and pattern detection, semantic features. The GNet layers occur later in the CNN, these filters are believed to learn more global features, that is combinations of basic shapes and edges. Following training, GNet remains static while the SNet is updated every 20 frames according to an adaption rule and a discrimination rule. The output of the GNet layer is fed to a distracter detector, this evaluates the confidence of the GNet prediction and if it falls below a threshold, it is assumed a

© Journal of Medical Artificial Intelligence. All rights reserved.

distractor is present and the output of the SNet is used to make the final prediction, in the absence of a distractor the output of the GNet is preferred. In comparison with Li *et al.* (32) this algorithm was found to perform 3% better with a precision score of 88% and accuracy score of 85%.

Soran *et al.* (33) use a 14-layer CNN to estimate the severity of spinal muscular atrophy, SMA type 1, in infants. Seventy, two-minute videos, taken at 30fps, of 10 subjects, up to 26 months of age, lying down and behaving naturally, were used to train the CNN. As CNN's require significant levels of data for training a sliding window approach was used to create multiple videos from the original set. The sliding window was set to step at 50 frame intervals and produce 30 second video snippets for training and testing. The issue of limb occlusion was resolved by ignoring sequences of data when this occurred; it is assumed this was done as a manual selection as no mention of techniques to identify occluded limbs is specified in the paper. The CNN was able to estimate the severity of SMA type-1 condition to and average error rate of less than 10%. This was a good result considering the relatively small amount of data available for training.

Moharir *et al.* (34) investigated the use of two different deep learning algorithms, AlexNet and GooLeNet, as classifiers, both these networks are CNNs, although their layer architecture differs considerably. The networks were trained on audio recordings, of 1 second duration, from children who had suffered asphyxia at birth and well as children who had normal births. As a note the abnormal to normal ratio of the training data was 2:1, compared to the natural occurrence rate for birth asphyxia of 0.86 per 1000 (35). The GooLeNet produced a classification accuracy rate of 94% compared to the AlexNet performance at 92%.

Shukla *et al.* (36) also used a pre-trained CNN, AlexNet, to classify images of subjects and detect a range of neurological conditions. The classifier was trained to classify images into six groups, autism, foetal alcohol syndrome, Down syndrome, intellectual disability and cerebral palsy. A pre-trained AlexNet, trained on the Labelled Faces in the Wild, LFW data set, was used as an encoder to automatically encode key features from the image to be classified. As the AlexNet CNN is not used to classify the images, a separate classifier, in this case a SVM was used. The training of this system involved presenting the images to the AlexNet, then taking the AlexNet outputs at layer fc7 and feeding these as inputs to the SVM, the SVM weights were then adjusted to obtain the desired classification; that is the AlexNet remained unchanged by the training process. In the deep neural network literature this technique is referred to as transfer learning (37).

Subjects in the Shukla *et al.* (36) study were from three age groups, 0 to 6 years, 6 to 12 years and 12 years and above. The classifier SVM was trained on a test set of 1,196 normal and 1,196 abnormal examples, again training on an elevated normal to abnormal ratio. The classifier was tested both against a test set and human experts; the results of these tests indicated the classifier performing at least at human level.

Table 1 summerises the results of the studies reviewed. The table shows that the current best practice baby movement classification systems produce accuracies between 85% and 90% with specificity and sensitivities at similar levels. By comparison expert classifiers achieve accuracies around 98% (4) indicating the available scope of improvement for this project.

Data source

A common problem for projects looking at automating the detection of cerebral palsy is the paucity of available data for training and testing, Zhang and Suganthan (38) tabulate 20 studies with an average of 80 CP and 135 Normal subjects. The Cerebral Palsy Alliance research foundation is currently running a data collection program. The parents of children under 6 months of age, who have cerebral palsy or are at high-risk of having cerebral palsy, are invited to participate. The participant's children undergo a number of assessments including a video recording of the child's movements. The video is shot on a constrained background and consists of the child lying calmly while being filmed to replicate the GMA assessment. Background shadings, blankets and mats, vary between videos, as videos are shot in the home environment. The recording application provides a shape template on the viewing screen to ensure scaling between videos is consistent. The videos are currently being correlated and labeled by the Cerebral Palsy Alliance, (CPA), research foundation; to date over 500 videos have been labeled. This resource is being made available to this project and represents a significant increase in the available data for a CP assessment project compared with prior projects.

On average a practitioner of Prechtl's General Movements assessment can make an initial assessment within 20 seconds of viewing the child's movements with a sensitivity and specificity of 98% and 91% (4). Using this assumption, the data set has from CPA been expanded by editing the 3 minutes videos into 20 second segments. Fidgety movements do not occur continuously, so 5–20 second segments were constructed at 20 second intervals, for a total data set of

© Journal of Medical Artificial Intelligence. All rights reserved.

Table 1 Comparison of baby movement classification algorithm performance

Study/method	Detecting	Vision/sensors	Sensitivity	Specifity	Accuracy
Rahamati (20), SV and hand features	FM's	2d video, depth channel, patches on limbs	86%	92%	83%
Ansari, Roy and Dogra (27)					
Skeleton features + HMM	Limb position during exercise	2d video	No report	No report	64%
Bag-of-words features + LSTM-RNN			No report	No report	65%
Bag-of-words features + HMM (one pass)			No report	No report	79%
Bag-of-words features + HMM (two-pass)			No report	No report	84%
Kahn, Helsper, Farid and Grezgorzek (28)					
Pre-processing Algorithm 1 and SV	Limb position during exercise	2d video, depth channel	86.01%	80.93%	83.29%
Pre-processing Algorithm 2 and SV			94.23%	88.08%	91.03%
Pre-processing Algorithm 3 and SV			98.20%	95.71%	97.00%
Stahl *et al.* (22), SV resting saliency of optical flow features					
Absolute motion distance	GM's	2d video	76.70%	95.10%	91.70%
Relative frequency			85.30%	95.50%	93.70%
Wavelet coefficient			56.00%	90.70%	84.40%

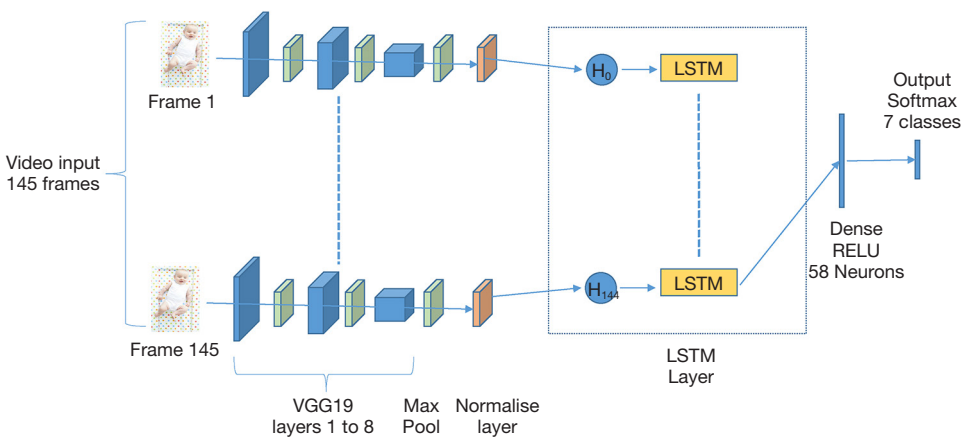

Figure 1 Pretrained VGG19 to LSTM—simple model.

2,445 video segments. To reduce the data per input video the 20 second segments were sampled 5 frame intervals, producing 145 frames per segmented video.

Initial approach

As an initial step a simple model has been constructed using a transfer learning approach. The model was constructed using the Keras VGG19 model, trained on the 1,000 classes of the ImageNet database (39). Image features were taken from Layer 8 of VGG19, passed through a max pooling layer and normalized (40) before being inputting to an LSTM layer for classification of the image sequence, see *Figure 1*.

© Journal of Medical Artificial Intelligence. All rights reserved.

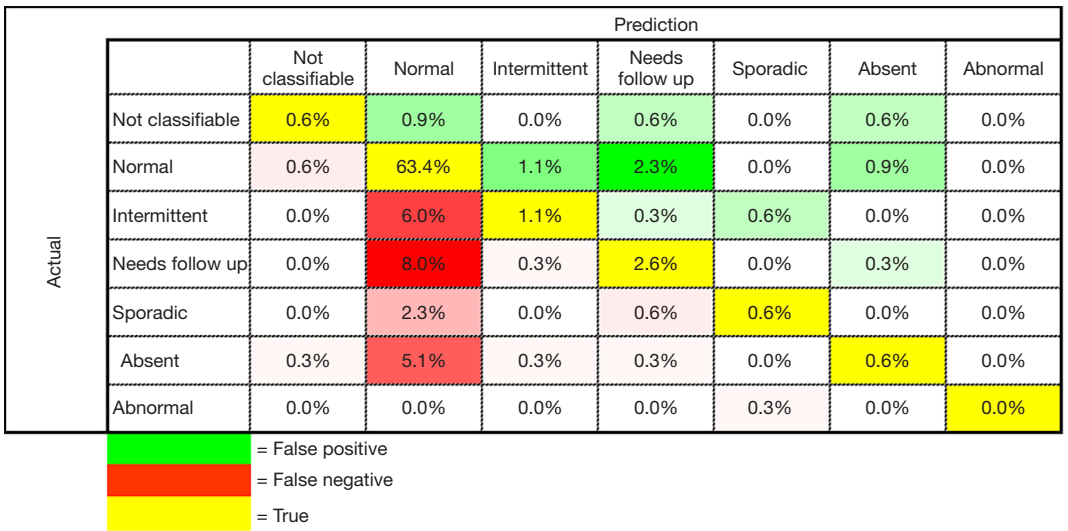

Figure 2 Confusion matrix.

Table 2 Data class distribution

Classification Class	% samples	Normal/Abnormal Dist
Normal	68.5%	
Intermittent	7.4%	75.9%
Not classifiable	3.1%	
Needs follow up	9.6%	
Sporadic	3.9%	
Absent	7.0%	
Abnormal	0.6%	24.1%

Initial results

A 10-fold cross validation strategy with an 80% training 20% test/holdout data split was adopted for training the model. The ADAM optimizer was adopted for training with a fixed learning rate of 0.001 and an early stopping strategy that produced average training runs of 120 epochs.

The first trained model produced an overall classification accuracy of 65.1% with a standard deviation of +/- 1%. Examination of the confusion matrix (*Figure 2*) shows that the model classifies Normal videos with great confidence, but struggles with intermittent and needing follow up, classes. Classification of these border line classes is also a difficult task for experts.

Discussion

The natural occurrence rate of CP in the data set, under 6 months of age at high-risk, is approximately 15% (M Fahey, 2019, personal communication). Such a ratio normal to abnormal, 85%:15%, represents a challenge for DNN training as the natural distribution is unbalanced. The data set reasonably reflects the expected natural distribution, see *Table 2* below.

In medical diagnosis sensitivity is preferred to specificity, i.e., the consequence of false negatives outweighs the consequences of false positives, at present the confusion matrices indicates that the model tends to sensitivity (50.8%) over specificity (27.4%), making it unsuitable in its current state, and performing poorly compared to the systems previously reviewed, sensitivity (>80%) and specificity (>90%).

These results are not unexpected as the project is in the early stages and with further work a more robust result is expected.

Proposed program

Moving forward the proposed program is to continue to explore the use of transfer learning, to pre-process the video frames to detect relevant features. An empirical examination of the convolutional layers of the VGG19, ResNet and Inception

© Journal of Medical Artificial Intelligence. All rights reserved.

models will be made to identify the most appropriate layers for sourcing the input data to the LSTM. A particular focus will be on hand and digit movement and detection and identifying the layers that match this level of resolution.

How much video is required before an LSTM can identify the presence of CP is a key research question. It is hypothesized that as a minimum the videos should be no shorter than required for an "expert" to make a positive identification. A test program will be run to establish this minimum length using the training data and a number of CP experts to evaluate a number of videos of varying lengths. Having established the minimum length required, the training set will be edited to produce multiple videos of similar lengths, using off sett frames to further augment the available data for training and testing of the system. The use of background subtraction to highlight motion in the input images, will be explored.

When making a diagnosis a key source of information is the frequency of occurrence, how likely a particular event or condition is dictates how much weight we assign to its co-occurring data. The review highlights that frequency of occurrence is information that is typically withheld from the trained classifier, when the classifier is used to assess a condition and there is minimal available training data. In typical cases the ratio of normal to abnormal conditions is near 50:50, while the natural occurrence rate tends to be closer to 1,000:1. The naïve argument in favour of these low ratios is the observation that we learn through repetition and the greater our exposure to a family of problems and their solutions, the better we become at solving these problems, this argument is extended to machine learning algorithms by analogy.

The analogy ignores the fact that unlike a learning algorithm, we are constantly updating our learning samples, hence we tend to experience higher ratios of normal/abnormal cases despite our best intentions otherwise. The problem is not un-recognized in machine learning, and is typically referred to as over fitting (41). An over fitting algorithm will recognise its test set with 100% accuracy and yet fail to generalise to previously unseen data. This behaviour is analogous to the human trait of a confirmation bias Wason as cited in (42) demonstrated this trait by showing subjects a sequence of three numbers and asking them to derive the rule that generated the numbers; 80% of the subjects derived the wrong rule and proceed to seek confirming evidence for their rule rather than attempting to falsify their rule. This results has since been replicated multiple times and is linked to a sense of self-esteem, as evidenced by tendency to select friends who hold similar views and beliefs (42). By making the naïve assumption, that more of the same is better, we may be in advertency biasing the network in a similar style. A further research question will be to explore the effects of normal/abnormal ratios, in the training data, on the chosen classifier.

Acknowledgments

Funding: None.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jmai.2019.06.02). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Brown RC, Lockwood AH, Sonawane BR. Neurodegenerative diseases: an overview of environmental risk factors. Environ Health Perspect 2005;113:1250.
2. Canavese F, Deslandes J. Orthopedic Management of Children with Cerebral Palsy: A Comprehensive Approach. Nova Science Publishers, Incorporated; 2015.
3. Hadders-Algra M. Early diagnosis and early intervention in cerebral palsy. Front Neurol 2014;5:185.
4. Herskind A, Greisen G, Nielsen JB. Early identification and intervention in cerebral palsy. Dev Med Child Neurol 2015;57:29-36.
5. The Australian Cerebral Palsy Register Group, Badawi N, Balde I, et al. Australian Cerebral Palsy Register Report

© Journal of Medical Artificial Intelligence. All rights reserved.

2016. 2016. doi: 10.13140/RG.2.2.36565.42726.
6. Shih STF, Tonmukayakul U, Imms C, et al. Economic evaluation and cost of interventions for cerebral palsy: a systematic review. Dev Med Child Neurol 2018;60:543-58.
7. McIntyre S, Morgan C, Walker K, et al. Cerebral palsy—don't delay. Dev Disabil Res Rev 2011;17:114-29.
8. Novak I. Evidence-based diagnosis, health care, and rehabilitation for children with cerebral palsy. J Child Neurol 2014;29:1141-56.
9. Granild-Jensen JB, Rackauskaite G, Flachs EM, et al. Predictors for early diagnosis of cerebral palsy from national registry data. Dev Med Child Neurol 2015;57:931-5.
10. Morgan C, Novak I, Dale RC, et al. Optimising motor learning in infants at high risk of cerebral palsy: a pilot study. BMC Pediatr 2015;15:30.
11. Blauw-Hospers CH, Hadders-Algra M. A systematic review of the effects of early intervention on motor development. Dev Med Child Neurol 2005;47:421-32.
12. Fairhurst C. Cerebral palsy: the whys and hows. Arch Dis Child Educ Pract Ed 2012;97:122-31.
13. Bosanquet M, Copeland L, Ware R, et al. A systematic review of tests to predict cerebral palsy in young children. Dev Med Child Neurol 2013;55:418-26.
14. Aisen ML, Kerkovich D, Mast J, et al. Cerebral palsy: clinical care and neurological rehabilitation. Lancet Neurol 2011;10:844-52.
15. Hubermann L, Boychuck Z, Shevell M, et al. Age at referral of children for initial diagnosis of cerebral palsy and rehabilitation: current practices. J Child Neurol 2016;31:364-9.
16. Kwong AK, Fitzgerald TL, Doyle LW, et al. Predictive validity of spontaneous early infant movement for later cerebral palsy: a systematic review. Dev Med Child Neurol 2018;60:480-9.
17. Einspieler C, Marschik PB, Pansy J, et al. The general movement optimality score: a detailed assessment of general movements during preterm and term age. Dev Med Child Neurol 2016;58:361-8.
18. Marcroft C, Khan A, Embleton ND, et al. Movement Recognition Technology as a Method of Assessing Spontaneous General Movements in High Risk Infants. Front Neurol 2015;5:284.
19. Blanco N, O'Hara LM, Robinson GL, et al. Health care worker perceptions toward computerized clinical decision support tools for Clostridium difficile infection reduction: A qualitative study at 2 hospitals. Am J Infect Control 2018;46:1160-6.
20. Rahmati H, Martens H, Aamo OM, et al. Frequency Analysis and Feature Reduction Method for Prediction of Cerebral Palsy in Young Infants. IEEE Trans Neural Syst Rehabil Eng 2016;24:1225-34.
21. Rahmati H, Dragon R, Aamo OM, et al., editors. Motion segmentation with weak labeling priors. German Conference on Pattern Recognition; 2014: Springer.
22. Stahl A, Schellewald C, Stavdahl O, et al. An optical flow-based method to predict infantile cerebral palsy. IEEE Trans Neural Syst Rehabil Eng 2012;20:605-14.
23. Friedman H, Soloveichick M, Barak S, et al. Neuroplasticity in Young Age: Computer-Based Early Neurodevelopment Classifier. 2018. Available online: https://www.intechopen.com/books/neuroplasticity-insights-of-neural-reorganization/neuroplasticity-in-young-age-computer-based-early-neurodevelopment-classifier
24. Garcia-Garcia A, Orts-Escolano S, Oprea S, et al. A survey on deep learning techniques for image and video semantic segmentation. Appl Soft Comput 2018;70:41-65.
25. Menze M, Geiger A, editors. Object scene flow for autonomous vehicles. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015.
26. Veeriah V, Zhuang N, Qi G-J, editors. Differential recurrent neural networks for action recognition. Proceedings of the IEEE international conference on computer vision; 2015.
27. Ansari AF, Roy PP, Dogra DP. Exercise classification and event segmentation in Hammersmith Infant Neurological Examination videos. Machine Vision and Applications 2018;29:233-45.
28. Khan MH, Helsper J, Farid MS, et al. A computer vision-based system for monitoring Vojta therapy. Int J Med Inform 2018;113:85-95.
29. Khan MH, Helsper J, Boukhers Z, et al., editors. Automatic recognition of movement patterns in the vojta-therapy using RGB-D data. Image Processing (ICIP), 2016 IEEE International Conference on; 2016: IEEE.
30. Li H, Li Y, Porikli F, editors. Robust online visual tracking with a single convolutional neural network. Asian Conference on Computer Vision; 2014: Springer.
31. Wang L, Ouyang W, Wang X, et al., editors. Visual tracking with fully convolutional networks. Proceedings of the IEEE international conference on computer vision; 2015.
32. Li Y, Lan C, Xing J, et al., editors. Online human action detection using joint classification-regression recurrent neural networks. European Conference on Computer

© Journal of Medical Artificial Intelligence. All rights reserved.

Vision; 2016: Springer.
33. Soran B, Lowes L, Steele KM, editors. Evaluation of Infants with Spinal Muscular Atrophy Type-I Using Convolutional Neural Networks. European Conference on Computer Vision; 2016: Springer.
34. Moharir M, Sachin M, Nagaraj R, et al., editors. Identification of asphyxia in newborns using gpu for deep learning. Convergence in Technology (I2CT), 2017 2nd International Conference for; 2017: IEEE.
35. Pierrat V, Haouari N, Liska A, et al. Prevalence, causes, and outcome at 2 years of age of newborn encephalopathy: population based study. Arch Dis Child Fetal Neonatal Ed 2005;90:F257-61.
36. Shukla P, Gupta T, Saini A, et al., editors. A deep learning frame-work for recognizing developmental disorders. Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on; 2017: IEEE.
37. Manaswi NK. Deep Learning with Applications Using Python. 2018.
38. Zhang L, Suganthan P. A Survey of Randomized Algorithms for Training Neural Networks. Inf Sci 2016;364-365:146-55.
39. He K, Zhang X, Ren S, et al., editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
40. Laurent C, Pereyra G, Brakel P, et al., editors. Batch normalized recurrent neural networks. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2016: IEEE.
41. Kubat M. An introduction to machine learning. Springer; 2015.
42. Myers DG, Myers RD. Social Psychology. McGraw-Hill Australia; 2013.

doi: 10.21037/jmai.2019.06.02
Cite this article as: Schmidt W, Regan M, Fahey M, Paplinski A. General movement assessment by machine learning: why is it so difficult? J Med Artif Intell 2019;2:15.

© Journal of Medical Artificial Intelligence. All rights reserved.