medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Title: Automated Leg Dystonia Diagnosis in Children with Cerebral Palsy Using Machine Learning

Authors: Rishabh Bajpai[1], Alyssa Rust[1], Emma Lott[1], Susie Kim[1], Sushma Gandham[1], Keerthana Chintalapati[1], Joanna Blackburn[2], Rose Gelineau-Morel[3], Michael C. Kruer[4], Dararat Mingbundersuk[5], Jennifer O'Malley[6], Laura Tochen[7], Jeff Waugh[8], Steve Wu[9], Timothy Feyma[10], Bhooma R. Aravamuthan[1]*

Affiliations:

[1] Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA
[2] Department of Pediatrics, Northwestern University, Chicago, IL
[3] Department of Pediatrics, University of Missouri-Kansas City School of Medicine, Kansas City, MO, USA
[4] Department of Child Health, University of Arizona College of Medicine Phoenix, Phoenix, AZ, USA
[5] Department of Neurology, Seattle Children's Hospital, University of Washington, Seattle, WA, USA
[6] Department of Neurology & Neurological Sciences, Stanford University, Palo Alto, CA, USA
[7] Department of Neurology and Rehabilitation Medicine, George Washington School of Medicine and Health Sciences, Washington DC, USA
[8] Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, USA
[9] Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA
[10] Department of Neurology, Gillette Children's Specialty Healthcare, St Paul, MN, USA.

*Corresponding author.
Division of Pediatric Neurology, Department of Neurology
Washington University School of Medicine
660 South Euclid Avenue, Campus Box 8111
St. Louis MO 63110-1093
Ph: 314-454-6120
Email: aravamuthanb@wustl.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Key Points

Question: Can fully automated machine learning-based software improve upon routine clinical rates of leg dystonia diagnosis in children with cerebral palsy?

Findings: Open-source software was developed that utilizes machine learning models to detect leg dystonia in clinically acquired videos of children with cerebral palsy from two different clinical centers, achieving higher sensitivity, specificity, positive-predictive value, negative-predictive value, and accuracy for dystonia diagnosis than has been demonstrated for routine clinical care.

Meaning: This software could facilitate screening children with cerebral palsy for dystonia, a common, debilitating, and under-diagnosed movement condition that requires accurate diagnosis for tailored treatment.

Abstract

Importance: Dystonia in cerebral palsy (CP) is common and debilitating but under-diagnosed, particularly in the legs. Only 13% of children with CP and leg dystonia have that dystonia diagnosed in clinic. Yet, early and accurate dystonia diagnosis is critical for appropriate treatment selection.

Objective: To facilitate automated leg dystonia diagnosis in CP, our objective was to develop open-source software that used machine learning (ML) models and clinically available video data.

Design, Setting, and Participants: In this cross-sectional study, videos of children with CP (age 5 and older) doing a seated alternating hand open-close task were acquired between 1/1/2020 and 12/14/23 from two tertiary care CP centers (193 videos from Center 1, 30 videos from 2). These videos were assessed for leg dystonia by 8 pediatric movement disorders physicians (gold-standard diagnosis by expert consensus). Movement features they cited when identifying dystonia were extracted from their consensus-building discussions using conventional content analysis and were then quantified from the videos leveraging our open-source software. Quantified features from 163 videos from Center 1 were used to train ML models (4664 models based on 11 ML algorithms) which were then tested on 30 videos from Center 1 and 30 videos from Center 2.

Main Outcome(s) and Measure(s): Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy of ML models were computed as compared to gold-standard leg dystonia diagnosis by expert consensus.

Results: Expert-cited quantifiable movements associated with dystonia diagnosis were used to identify 63 kinematics features to extract from videos. Feature importance analysis showed that variance features had higher importance than amplitude features. ML models achieved 84% sensitivity, 84% specificity, 84% PPV, 85% NPV, and 84% accuracy for identifying leg dystonia in testing videos from Center 1 and 100% sensitivity, 67% specificity, 75% PPV, 100% NPV, and 83% accuracy for identifying leg dystonia in videos from Center 2. However, for videos where at least 60% of experts agreed regarding the presence or absence of dystonia (n=23/30 from Center 2), ML models were 100% accurate for identifying leg dystonia.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Conclusions/Significance: We have developed open-source software that can use clinically acquired videos to identify leg dystonia in children with CP. These methods may be useful for facilitating clinical dystonia diagnosis in children with CP.

Introduction

Dystonia is a debilitating movement disorder formally defined as "sustained or intermittent muscle contractions causing abnormal, often repetitive movements or postures".[1] The most common condition associated with dystonia in children is cerebral palsy (CP), which affects 2-4 of every 1000 people globally.[2–4] Early and accurate dystonia diagnosis in people with CP is crucial for many reasons: treatments for dystonia are distinct from treatments for other movement conditions like spasticity,[5,6] there is evidence that dystonia-specific treatments work best if administered early,[7] and dystonia is a relative contraindication to some surgical procedures commonly done in young children with CP.[8–10] Current clinical practices for diagnosing dystonia rely heavily on the expertise of trained clinicians (e.g. pediatric movement disorders specialists). The gold standard for dystonia diagnosis is the consensus opinion of these experts as they review video or in-person assessments of a person's movements. However, these experts are in short supply and not available at all centers.[11] Furthermore, dystonia is variable by definition,[1,12] making clinical diagnosis at one time point by one clinician difficult. For example, a clinic visit with an expert clinician has 13% sensitivity, 11% specificity, 18% positive predictive value (PPV), 8% negative predictive value (NPV), and 12% accuracy for leg dystonia diagnosis in a young person with CP when compared to gold standard consensus-based expert assessment.[13] Clinical dystonia underdiagnosis in the legs is particularly notable given that dystonia most commonly affects the legs in people with CP.[14] Clearly, there is a dire need to facilitate clinical dystonia diagnosis.

Though there are many scales to guide dystonia severity assessment after it has been accurately identified, the only clinical tool available to facilitate dystonia diagnosis in people with CP is the Hypertonia Assessment Tool (HAT).[15,16] The HAT provides a valuable framework to distinguish dystonia from spasticity and rigidity. However, the items governing dystonia diagnosis have the lowest interrater reliability and the movements that must be recognized to identify dystonia when using the HAT are subtle.[15,16] Therefore, for the HAT to work optimally as a screening diagnostic tool, it must also be employed by well-trained clinicians.

Noting that dystonia diagnosis heavily relies on visual assessment, recent advances in video-based pose estimation can facilitate dystonia diagnosis. Open-source tools for identifying body coordinates in clinically-acquired videos (e.g. DeepLabCut[17] and OpenPose[18]) can facilitate movement quantification. Furthermore, rigorous qualitative analysis of the language experts use to describe dystonia can help determine the most valuable movement features to quantify.[19] Using these methods, we recently demonstrated that leg adduction variability and leg adduction amplitude correlate with expert assessments of leg dystonia severity in people with CP during gait.[20] However, it is unclear what features govern experts' leg dystonia assessments in people with CP as they perform other tasks (noting that not all people with CP can walk). We hypothesize that expert-cited features can be quantified in clinically acquired videos and used to train machine learning (ML) models to facilitate leg dystonia diagnosis.

In this study, we asked eight pediatric movement disorders physicians to assess for leg dystonia in videos of children with CP doing a seated motor task. Using their consensus-building discussions, we determined the movement features they most often cited when diagnosing dystonia. We then quantified these expert-cited features in the same videos using body coordinates extracted with

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

OpenPose[18] and used these quantifications to train, validate, and test dystonia diagnostic ML models. We hypothesized that this approach could yield an automated dystonia diagnostic model that outperforms current clinical leg dystonia diagnostic rates in children with CP (13% sensitivity, 11% specificity, 12% accuracy)[13] when compared to gold-standard expert consensus.

Methods

Standard protocol approvals, registrations, and patient consents

This study received approval from the Washington University in St. Louis Institutional Review Board on 6/5/23 (Approval Number: 202102101).

Subjects, motor task, and video acquisition

Videos of young people with CP are acquired during routine clinical care at Gillette Children's Hospital (henceforth called Center 1) and St. Louis Children's Hospital (henceforth called Center 2). Both centers record people with CP performing an alternating hand open/close task. We have previously established that this task is particularly valuable for distinguishing the presence vs. absence of dystonia.[21] Videos of this task at both sites are of the person sitting in a chair facing forward, resting their non-dominant hand on the ipsilateral thigh, and then raising and alternately opening and closing/fisting their dominant hand as quickly as possible for approximately 5 seconds. This task is then repeated with the non-dominant hand. At Center 1, videos were recorded with a Panasonic AG-AC160AP camera at 1920 x 1080 pixel resolution at 30 frames per second. At Center 2, videos are recorded with a Google Pixel 5 smartphone camera at 3840 x 2160 pixel resolution at 60 frames per second rate.

Videos for this study were taken during routine clinical care between 1/1/2020 and 12/30/2021 at Center 1 and between 4/20/2023 to 12/14/2023 at Center 2. To be included, subjects had have a clinician-confirmed diagnosis of CP according to the 2006 consensus diagnostic criteria[22] and had to be at least 5 years old, with this age cutoff chosen to increase the likelihood that the child would developmentally have the cognitive ability and attention span to attempt the motor task in full. Exclusion criteria were the full body not being visible in the video or the person being unable to complete the task during the video recording. Videos from Center 1 were all videos meeting inclusion and exclusion criteria recorded during the study period. Videos from Center 2 were 30 videos chosen from all eligible videos to have an equal distribution of subjects who had been clinically determined to have dystonia and not have dystonia, as determined by each person's treating movement disorders clinician using the Hypertonia Assessment Tool. [15,16]

Consensus-based expert video assessment

To ensure subject anonymity, the faces of all participants in the videos were blurred using ShotCut (Meltytech, LLC) and custom-written software.[23] A subset of 24 videos was reviewed in consensus by 8 pediatric movement disorders physician experts from different institutions across the United States (JB, RGM, MK, DM, JOM, LT, JW, SW). Leg dystonia severity was rated by each expert using the Global Dystonia Rating Scale (GDRS),[24] a 10-point Likert scale (0 - no dystonia, 10 - severe dystonia) with high inter-rater reliability for assessing leg dystonia severity in people with CP.[20] For computing the GDRS for leg dystonia severity, experts rated the bilateral proximal legs (including hips) and bilateral distal legs (including feet). Combining these ratings resulted in a GDRS score range of 0-40.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

To build consensus, experts participated in three 1-hour Zoom sessions (Zoom Video Communications, Inc.) to discuss leg dystonia severity in each of 24 videos. After rating leg dystonia severity using the GDRS in these consensus-building sessions, the remaining videos from Center 1 by at least two of the eight experts, with scores entered via a REDCap survey. For any videos where the GDRS scores differed by more than 6 out of 40 possible points between two experts, the entire group of 8 experts re-assessed the video. Videos from Center 2 were each rated by 7 experts (JB, RGM, DM, JOM, LT, JW, SW). The final leg dystonia severity assessment for each video was calculated as the median GDRS across all expert reviewers with GDRS<1 indicating leg dystonia was absent and GDRS≥1 indicating that leg dystonia was present.

Qualitative analysis of consensus-building discussions

Consensus-building sessions were recorded and transcribed. Transcripts were analyzed using conventional content analysis[25] to determine the salient ideas, or codes, stated by experts when assessing videos for leg dystonia. Transcripts were coded independently by two coders (SG, KC), who met to resolve any coding discrepancies by consensus and then consolidated their codes into a single code book. Potentially quantifiable movements (e.g. “foot inversion”) were distinguished from non-quantifiable codes (e.g. “posturing”) and used to calculate kinematics analogues from videos for training ML models.

Extraction of leg coordinates

We used custom-written software[23] leveraging OpenPose[18] to estimate leg coordinates that were consistently visualizable in the videos: the midpoint of the patella, the midpoint between the medial and lateral malleolus, and the tip of the first toe for both the right and left legs. To ensure labeling accuracy, the extracted coordinates were overlayed on the videos and visually examined.

Determining the features used to train dystonia diagnostic models

Two categories of leg kinematics were calculated: 1) the 2D leg coordinates (n=12, X and Y coordinates of the right and left knee, ankle, and first toe), and 2) calculated analogues of quantifiable movements cited by experts when assessing videos for dystonia. Since we previously demonstrated that variability and amplitude of leg adduction are cited by experts when assessing leg dystonia severity during gait and also correlate with their GDRS-based assessments of leg dystonia severity,[20] the variance, minimum, and maximum of all leg kinematics measures were used as the input features to train ML models for leg dystonia diagnosis.

Dystonia diagnostic model training, validation, and testing

A total of 4664 ML models were assessed, derived from 11 different ML algorithms: Logistic Regression,[26] Decision Tree,[27] Random Forest,[28] Gradient Boosting,[29] AdaBoost,[30] Support Vector Machine,[31] K-Nearest Neighbors,[32] Naive Bayes,[33] XGBoost,[34] Extra Trees,[35] and Bagging.[36] Videos from Center 1 were divided into subsets for training (2/3 of the dataset), validation (1/6 of the dataset), and testing (1/6 of the dataset), ensuring subsets had comparable GDRS distributions. To avoid bias from any single arbitrary division into subsets, ML models were trained, validated, and tested on five different subset groupings of the full video dataset, with each grouping having different training, validation, and testing subsets. Within each grouping, training and validation were repeated

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

five times using different subsets of videos for training and validation each time (5-fold cross-validation using the scikit-learn library) (Figure 1).

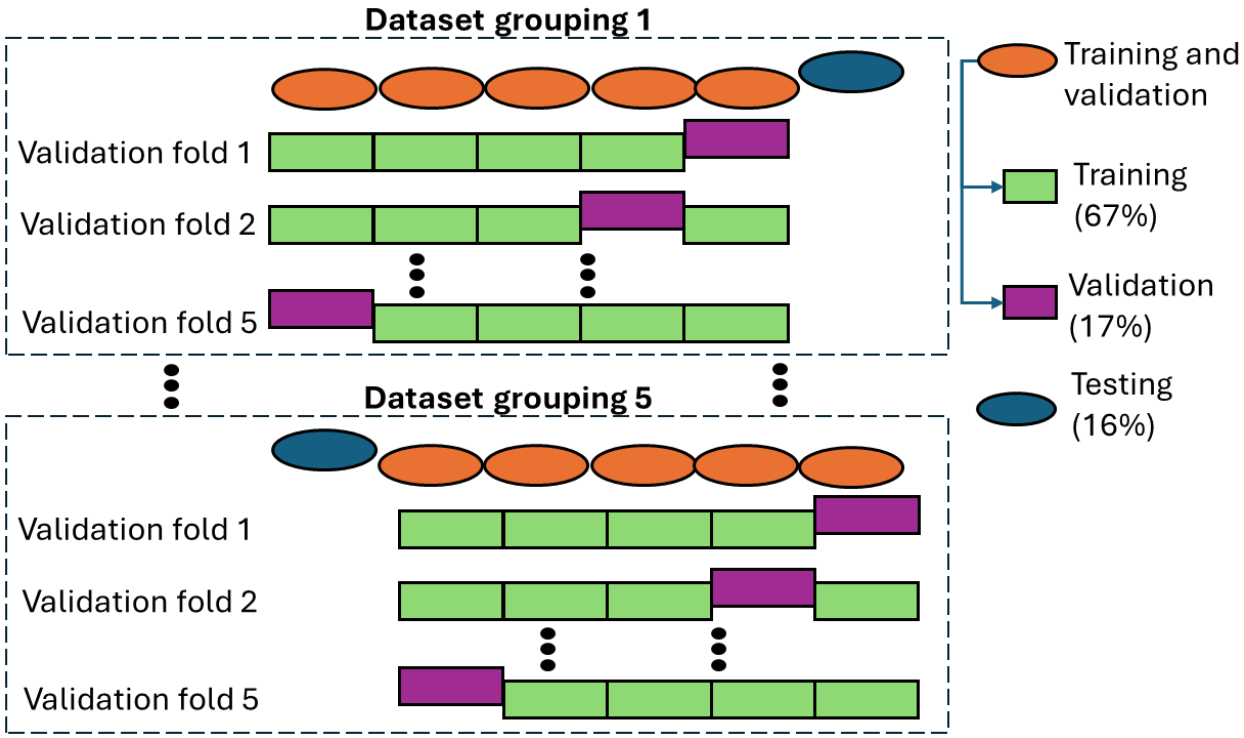

Figure 1. Dystonia diagnostic model training, validation, and testing. The full video dataset is divided into different subsets for training (2/3 of the dataset), validation (1/6 of the dataset), and testing (1/6 of the dataset) in each of 5 different dataset groupings of the full dataset. Within each dataset grouping, training and validation occurs 5 times, using different subsets for training and validation each time.

For each dataset grouping, a single ML model was selected for testing (out of the 4664 ML models used for training and validation) based on having the highest NPV and specificity averaged across all 5 validation folds. NPV and specificity were prioritized because the majority of people with CP have leg dystonia,[14] suggesting that a model's ability to detect the absence of dystonia may yield the best diagnostic model overall. To determine the optimal number of ML model features needed to maximize NPV and specificity, we first calculated feature importance within each dataset grouping using eight feature selection methods: Spearman's rho,[37] Kendall's tau,[38] ANOVA F-test/univariate feature selection,[39] Extra Trees Classifier,[35] Random Forest,[28] Maximum Relevance - Minimum Redundancy,[40] Sequential Feature Selection (forward),[41] and Recursive Feature Elimination.[42] Features were then ranked in descending order of feature importance for each grouping averaged across the eight feature selection methods, from 1 to N. Models were first trained on just the top ranked feature, then the top 2 ranked features, then the top 3 ranked features, and so on until the model was trained using all features). The model trained on the optimal number of features to maximize NPV and specificity was tested on the testing dataset for each of the five dataset groupings to determine the sensitivity, specificity, PPV, NPV, and accuracy of the model for leg dystonia diagnosis using videos from Center 1.

Of the N features used to train the ML models, we determined the features most relevant for leg dystonia diagnosis in two ways:

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1. We compared the distributions of feature importance rankings across variance, minimum, and maximum features using a one-way ANOVA.
2. We determined the training features shared by the ML models that maximized NPV and specificity across all five dataset groupings.

We created an average ensemble of the best performing models (highest NPV and specificity) from each of the five dataset groupings from Center 1 and tested this average ensemble using a separate dataset of the 30 videos from Center 2 selected to have comparable distributions of people with and without dystonia as per the treating clinician's assessment. Ensemble averaging[43] is a common method used aggregate the results of multiple ML models and is the simple average of each model's output. In this case, the dichotomous results of each of the five models for each subject (0 = leg dystonia absent, 1 = leg dystonia present) were averaged and rounded to the nearest integer value to determine the average ensemble determination.

Results

Subjects and demographics

At Center 1, 193 subjects met inclusion and exclusion criteria (age 7.2±3.1 years, 92.75% at Gross Motor Function Classification System Levels I-III, 91% with leg spasticity). Of these subjects, 54% (n=104) had an average GDRS≥1 and were thus designated as having leg dystonia (Supplementary eTable 1). Videos were 7.2±3.5 seconds in duration.

At Center 2, 30 selected subjects (age 12.2±4.4 years, 100% at Gross Motor Function Classification System Levels I-II, 100% with leg spasticity) were divided equally between those with leg dystonia (n=15, average GDRS >=1) and those without leg dystonia (n=15, GDRS<0) (Supplementary eTable 1). Videos were 51.5±3.2 seconds in duration.

Features used to train ML models

In addition to 2D leg coordinates (n=12 for X and Y coordinates of the left and right knee, ankle, and first toe), we also calculated kinematic analogues using these 2D leg coordinates to reflect features cited by experts when assessing dystonia.

Experts cited 15 potentially quantifiable movements a total of 140 times over 3 hours of consensus building discussion when assessing videos for dystonia (Figure 2).

These quantifiable movements were used to develop 11 kinematic analogues (Table 1). Analogues for three out of fifteen quantifiable movements cited by experts (namely, toe flexion, toe extension and "swan neck" toe posturing) were not calculated as this would all require reliable extraction of the coordinates of multiple toes, which could not be accurately and reliably obtained using OpenPose following our visual quality inspection.

To account for differences in subject size and subject-to-camera distances, the 12 leg coordinates and 11 calculated movement analogues (total n=23) were divided by the maximum knee-to-ankle distance from each video.[44]

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

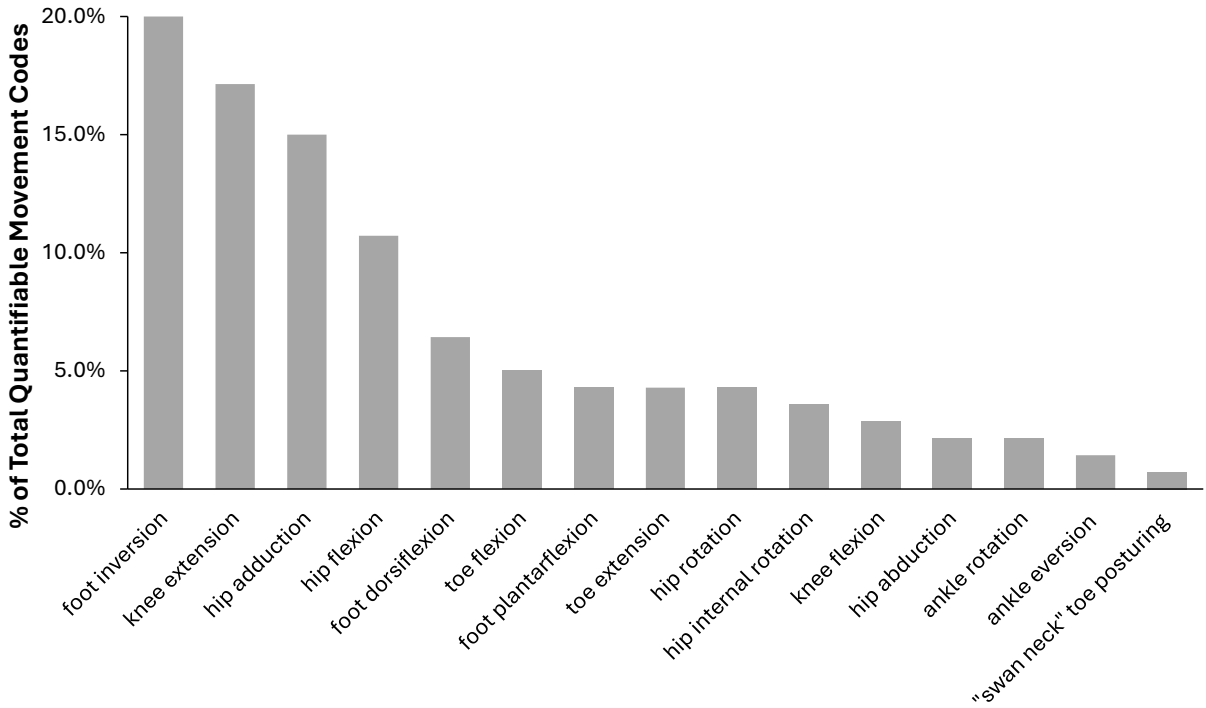

Figure 2. Frequencies of quantifiable movements cited by experts when diagnosing dystonia.

Quantifiable movements most frequently cited by experts	Calculated kinematic analogues	2D coordinate analogues
Foot inversion / ankle eversion/ ankle rotation	| Left Toe (X) – Left Knee (X) | | Left Toe (X) – Left Ankle (X) | | Right Toe (X) – Right Knee (X) | | Right Toe (X) – Right Ankle (X) | | Left Toe (X) – Right Toe (X) |	
Knee extension / knee flexion	| Left Toe (Y) – Left Knee (Y) | | Right Toe (Y) – Right Knee (Y) |	Left Toe (Y) Right Toe (Y)
Hip adduction / hip abduction	| Left Knee (X) – Right Knee (X) | | Left Ankle (X) – Right Ankle (X) |	
Hip flexion		Left Knee (Y) Right Knee (Y) Left Ankle (Y)

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

		Right Ankle (Y)
Foot dorsiflexion	| Left Toe (Y) – Left Ankle (Y) | | Right Toe (Y) – Right Ankle (Y) |	Left Toe (Y) Right Toe (Y)
Hip rotation/ hip internal rotation	| Left Toe (X) – Left Knee (X) | | Left Toe (X) – Left Knee (X) | | Left Toe (X) – Right Toe (X) |	

Table 1. Expert cited quantifiable movements, their calculated analogues, and 2D coordinate analogues. |a-b| represents absolute difference between 'a' and 'b'.

Comparison of feature importance rankings

The variance, minimum, and maximum of the coordinates and calculated analogues yielded 69 features used for training ML models. Their feature importance rankings averaged across all five dataset groupings are shown in Supplementary eTable 2. Variance features are ranked significantly higher than minimum and maximum features (one-way ANOVA, $p<0.005$, Figure 3).

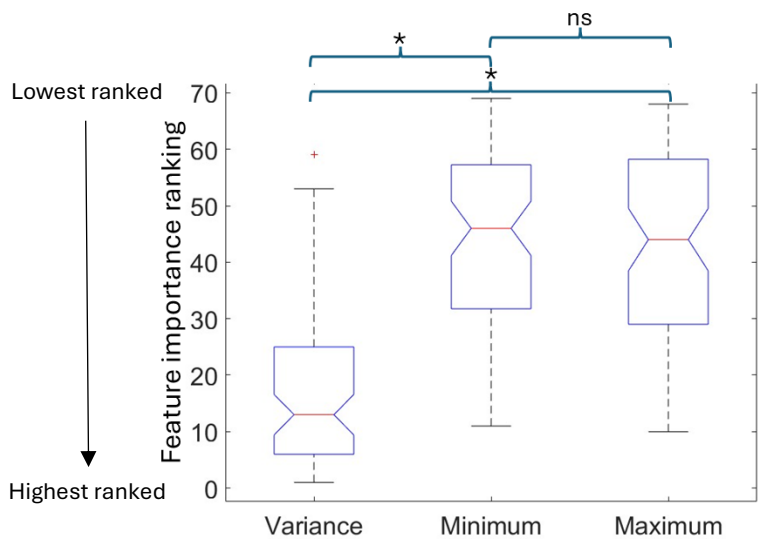

Figure 3. Comparison of the mean importance ranking of variance, minimum and maximum features. Features are ranked from 1 (highest importance) to 69 (lowest importance). The mean feature importance ranking for the variance features was significantly better than the mean rankings for minimum and maximum features (one-way ANOVA, *$p<0.005$). Whiskers indicate the interquartile range, the upper and lower edges of the boxes represent the 75th and the 25th percentile of the distribution, respectively, and the red middle line represents the median of the distribution.

Features shared by high performing ML models

Out of the eleven ML algorithms, the Naive Bayes[33] algorithm demonstrated the highest performance (highest specificity and NPV) in all 5 dataset groupings. These 5 ML models were trained on the

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

following numbers of top-ranked features: 38, 50, 31, 38, and 35 (Figure 4). Of these features, 25 were shared across all 5 ML models. Seventeen out of the twenty-five selected features were variance features, four were minimum features, and four were maximum features (Table 2). Across all 5 dataset groupings from Center 1, ML models obtained a mean sensitivity, specificity, PPV, NPV, and accuracy of 87%, 85%, 84%, 88% and 86%, respectively, during validation (Table 3) and 84%, 84%, 84%, 85% and 84%, respectively, during testing (Table 4).

Average ensemble model and testing on videos from Center 2

When tested on 30 videos from Center 2, the average ensemble model had sensitivity, specificity, PPV, NPV, and accuracy of 100%, 67%, 75%, 100% and 83%, respectively. Notably, there was disagreement between the 7 raters regarding whether leg dystonia was present or absent in each of these 30 videos: All 7 raters agreed regarding the presence or absence of leg dystonia in 16/30 videos, only 6/7 raters agreed for 7 videos, and only 4/7 raters agreed for 7 videos. When considering only the 23 out of 30 videos where more than 4/7 raters agreed regarding the presence or absence of leg dystonia, the average ensemble model always accurately predicted the majority rater responses (i.e. sensitivity, specificity, PPV, NPV, and accuracy for leg dystonia diagnosis was 100%).

Feature importance ranking	Variance feature	Feature importance ranking	Minimum feature	Feature importance ranking	Maximum feature
1	Variance |Right Toe(X) - Right Ankle (X)|	14	Minimum |Left Toe(Y) - Left Ankle (Y)|	16	Maximum |Left Toe(Y) - Left Ankle (Y)|
2	Variance |Left Toe(X) - Left Ankle (X)|	21	Minimum |Right Toe(Y) – Right Ankle (Y)|	17	Maximum |Right Toe(Y) – Right Ankle (Y)|
3	Variance Left Toe (X)	23	Minimum Inter Ankle (X)	22	Maximum Inter Toe(X)
4	Variance Left Ankle (Y)	25	Minimum Left Toe (Y)	24	Maximum Inter Ankle (X)
5	Variance Right Toe (Y)				
6	Variance Inter Ankle (X)				
7	Variance Left Toe (Y)				
8	Variance |Left Toe(Y) - Left Ankle (Y)|				
9	Variance Inter Toe(X)				
10	Variance |Right Toe(Y) – Right Ankle (Y)|				
11	Variance |Left Toe(Y) - Left Knee (Y)|				
12	Variance Right Ankle (Y)				
13	Variance Right Toe (X)				

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

15	Variance |Right Toe(Y) - Right Knee (Y)|				
18	Variance |Left Toe(X) – Left Knee (X)|				
19	Variance Right Ankle (X)				
20	Variance Left Ankle (X)				

Table 2. Features used to train all five optimal ML models across five dataset groupings.

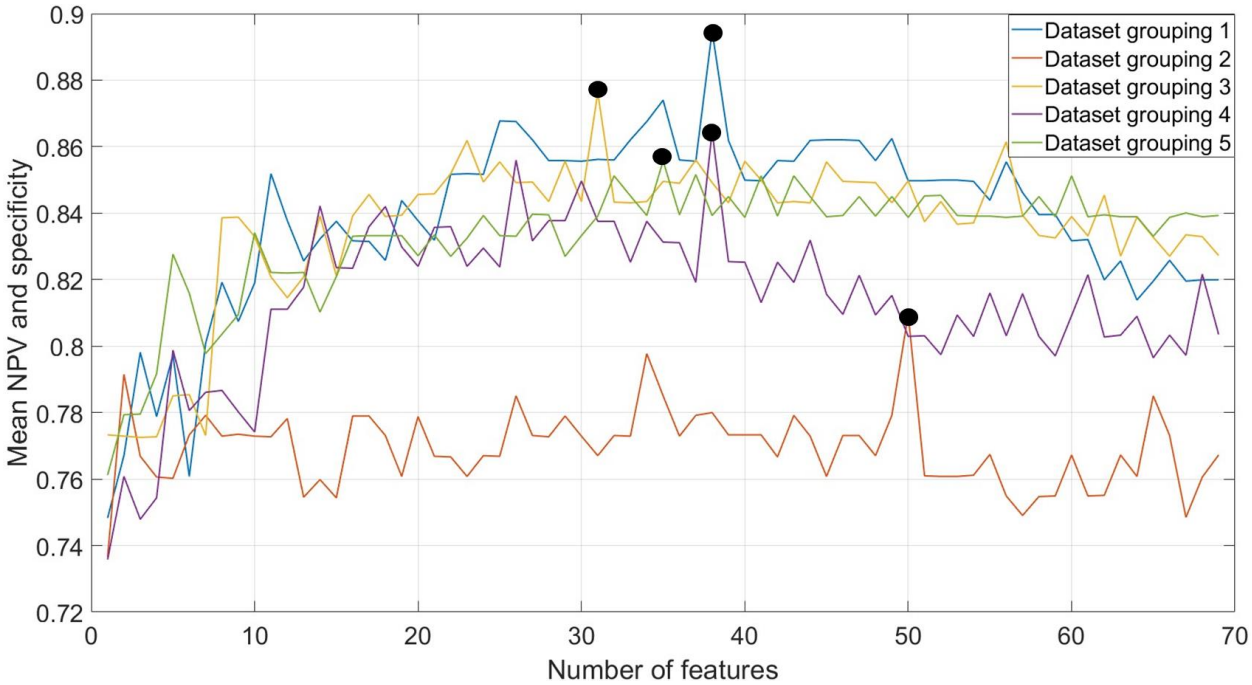

Figure 4. The mean of NPV and specificity of ML models when trained on different numbers of top-ranked features. The black points denote the best performing model for each of the five dataset groupings of the full dataset. The x-axis represents the number of features used for training and validation. For instance, a value of 50 on the x-axis indicates that the top 50 features, according to their importance, were used for training and validating the models. The y-axis represents the mean validation NPV and specificity of the best-performing ML model (out of 4664 models) when trained on the selected number of features.

Dataset Grouping	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	Optimal number of features
1	91	86	85	91	88	38

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2	81	81	80	82	81	50
3	86	90	88	88	88	31
4	86	87	86	87	87	38
5	91	80	80	91	85	35
Mean	**87**	**85**	**84**	**88**	**86**	**38.4**

Table 3. ML model performance during validation across all five dataset groupings from Center 1. The number of features used to train the model was selected to maximize NPV and specificity (Figure 4). NPV – negative predictive value; PPV- positive predictive value.

Model with highest NPV and specificity from each dataset grouping	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
1	80	93	92	82	87
2	100	80	83	100	90
3	93	87	88	93	90
4	80	73	75	79	77
5	67	87	83	72	77
Mean	**84**	**84**	**84**	**85**	**84**

Table 4. ML model performance during testing across all five dataset groupings from Center 1. NPV – negative predictive value; PPV- positive predictive value.

Average Ensemble Model	Sensitivity (%)	Specificity (%)	PPV (%)	NPV(%)	Accuracy(%)
All 30 videos	100	67	100	75	83
23 videos for which more than 4/7 experts agreed regarding presence/absence of leg dystonia	100	100	100	100	100

Table 5. ML model performance during testing on dataset from Center 2. NPV – negative predictive value; PPV- positive predictive value.

Software graphical user interface (GUI) development

An open-source software package (DxTonia)[23] was developed using 250 pre-defined and 2 custom libraries (details in eAppendix 1), 2 opensource software packages (OpenPose[18] and exiftool[45]) and

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the five best performing ML models described above. The software provides an interactive GUI to automate the process of leg dystonia detection from HOC videos. Its architecture is explained in Supplementary eFigure 2 and a demonstration video is available here: https://drive.google.com/drive/folders/1WGG4dfpLej4GRxtMJABAqe5nUVV-LftV?usp=sharing .

Key features of this software facilitating usability are:

1. Download and use without any installation.
2. Fully automated leg dystonia detection with a single click—no prior movement disorder expertise is required.
3. Fast diagnosis: The software takes approximately 5 times the video duration for first-time use and 0.5 times the video duration for subsequent uses.
4. Provides body coordinates in JSON, CSV, and Python variables, and dystonia ratings in JSON format for post-processing.
5. Open-source, allowing integration with other machine learning models.

Discussion

We developed an open-source software package (DxTonia) that uses ML models that identify leg dystonia in videos of children with CP performing a seated motor task with high sensitivity, specificity, PPV, NPV, and accuracy compared to gold-standard expert consensus assessment. Our results additionally suggest 25 leg kinematic features that may be most important for detecting leg dystonia and also suggest that variability of leg movements, more so than the amplitude of leg movements, may be most important for detecting leg dystonia. Finally, given low rates of leg dystonia diagnosis during routine clinical care, our hope is that software like what we have described here can help facilitate clinical screening for leg dystonia.

Limitations and the future work

The performance of ML models is heavily reliant on dataset size. With only 193 subjects' data, we couldn't fully leverage deep learning's potential for classification. Despite using 5-fold cross-validation and 5-fold testing, it is challenging to comment on the generalizability of the model's performance to datasets from different demographic populations, cameras, and environmental settings. While our study demonstrated encouraging results for the ML models' performance on an independent dataset from another center, we acknowledge that further validation with a broader and more diverse range of datasets would be beneficial in fully assessing its generalizability. Thus, future work will involve refining and testing the models across various clinical institutions and environmental conditions.

All stages of the proposed assessment, except for data collection are fully automated in the proposed software. Moving forward, we aim to develop an automated pipeline for automated data collection and secure local storage using smartphones, kinematics extraction via SecurePose[46] (automated body estimation software), and generation of leg dystonia assessment reports. Such a system would improve the accessibility of leg dystonia assessment.

Conclusion

We have developed an open-source software package (DxTonia) that has the potential to facilitate clinical leg dystonia assessment in children with CP. Future work will focus on broader validation of this software in larger populations across additional centers.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Albanese, A. *et al.* Phenomenology and classification of dystonia: a consensus update. *Mov. Disord.* **28**, 863–873 (2013).

2. McIntyre, S. *et al.* Global prevalence of cerebral palsy: A systematic analysis. *Dev. Med. Child Neurol.* **64**, 1494–1506 (2022).

3. Rice, J., Skuza, P., Baker, F., Russo, R. & Fehlings, D. Identification and measurement of dystonia in cerebral palsy. *Dev. Med. Child Neurol.* **59**, 1249–1255 (2017).

4. Steeves, T. D., Day, L., Dykeman, J., Jette, N. & Pringsheim, T. The prevalence of primary dystonia: A systematic review and meta-analysis. *Mov. Disord.* **27**, 1789–1796 (2012).

5. Fehlings, D. *et al.* Pharmacological and neurosurgical management of cerebral palsy and dystonia: Clinical practice guideline update. *Dev. Med. Child Neurol.* (2024) doi:10.1111/DMCN.15921.

6. Lott, E. *et al.* Physician Approaches to the Pharmacologic Treatment of Dystonia in Cerebral Palsy. *Pediatrics* **154**, (2024).

7. Lumsden, D. E. *et al.* Proportion of life lived with dystonia inversely correlates with response to pallidal deep brain stimulation in both primary and secondary childhood dystonia. *Dev. Med. Child Neurol.* **55**, 567–574 (2013).

8. Van De Pol, L. A. *et al.* Risk Factors for Dystonia after Selective Dorsal Rhizotomy in Nonwalking Children and Adolescents with Bilateral Spasticity. *Neuropediatrics* **49**, 44–50 (2018).

9. Gillespie, C. S. *et al.* The effect of GMFCS level, age, sex, and dystonia on multi-dimensional outcomes after selective dorsal rhizotomy: prospective observational study. *Childs. Nerv. Syst.* **37**, 1729–1740 (2021).

10. Blumetti, F. C. *et al.* Orthopaedic Surgery in Dystonic Cerebral Palsy. *J. Pediatr. Orthop.* **39**, 209–216 (2019).

11. Kahlon, S. *et al.* Emerging Subspecialties: Pediatric Movement Disorders Neurology. *Neurology* **102**, (2024).

12. Sanger, T. D., Delgado, M. R., Gaebler-Spira, D., Hallett, M. & Mink, J. W. Classification and Definition of Disorders Causing Hypertonia in Childhood. *Pediatrics* **111**, e89–e97 (2003).

13. Aravamuthan, B., Pearson, T. S., Chintalapati, K. & Ueda, K. Under-recognition of leg dystonia in people with cerebral palsy. *Ann. Child Neurol. Soc.* **1**, 162–167 (2023).

14. Rice, J., Skuza, P., Baker, F., ... R. R.-... M. & C. & 2017, undefined. Identification and measurement of dystonia in cerebral palsy. *Wiley Online Libr. Rice, P Skuza, F Bak. R Russo, D FehlingsDevelopmental Med. Child Neurol. 2017•Wiley Online Libr.* **59**, 1249–1255 (2017).

15. Jethwa, A. *et al.* Development of the Hypertonia Assessment Tool (HAT): a discriminative tool for hypertonia in children. *Wiley Online Libr. Jethwa, J Mink, C Macarthur, S Kn. T Fehlings, D FehlingsDevelopmental Med. Child Neurol. 2010•Wiley Online Libr.* **52**, (2010).

16. Knights, S. *et al.* Further evaluation of the scoring, reliability, and validity of the Hypertonia

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Assessment Tool (HAT). *journals.sagepub.comS Kn. N Datoo, A Kawamura, L Switzer, D FehlingsJournal Child Neurol. 2014•journals.sagepub.com* **29**, 500–504 (2014).

17. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci. 2018 219* **21**, 1281–1289 (2018).

18. Cao, Z., Hidalgo, G., Simon, T., Wei, S. E. & Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 172–186 (2018).

19. Aravamuthan, B. R. *et al.* Gait features of dystonia in cerebral palsy. *Dev. Med. Child Neurol.* **63**, 748–754 (2021).

20. Aravamuthan, B. R. *et al.* Determinants of gait dystonia severity in cerebral palsy. *Dev. Med. Child Neurol.* **65**, 968–977 (2023).

21. Gilbert, L. A. *et al.* Upper Extremity Dystonia Features in People With Spastic Cerebral Palsy. *Neurol. Clin. Pract.* **13**, (2023).

22. Rosenbaum, P., Paneth, N., Leviton, A., Goldstein, M. & Bax, M. A report: the definition and classification of cerebral palsy April 2006. *Dev. Med. Child Neurol.* **49**, 8–14 (2007).

23. DxTonia download | SourceForge.net. https://sourceforge.net/projects/dxtonia/.

24. Comella, C. L., Leurgans, S., Wuu, J., Glenn, S. T. & Chmura, T. Rating scales for dystonia: a multicenter assessment. *Mov. Disord.* **18**, 303–312 (2003).

25. Miles Matthew B, A. M. H. Qualitative Data Analysis. *SAGE Publ.* **1304**, 354 (1994).

26. Cramer, J. S. The Origins of Logistic Regression. *SSRN Electron. J.* (2002) doi:10.2139/SSRN.360300.

27. Fürnkranz, J. Decision Tree. *Encycl. Mach. Learn.* 263–267 (2011) doi:10.1007/978-0-387-30164-8_204.

28. Fawagreh, K., Gaber, M. M. & Elyan, E. Random forests: from early developments to recent advancements. *Syst. Sci. Control Eng. An Open Access J.* **2**, 602–609 (2014).

29. Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **54**, 1937–1967 (2021).

30. Schapire, R. E. Explaining adaboost. *Empir. Inference Festschrift Honor Vladimir N. Vapnik* 37–52 (2013) doi:10.1007/978-3-642-41136-6_5/TABLES/2.

31. Schölkopf, B. SVMs - A practical consequence of learning theory. *IEEE Intell. Syst. Their Appl.* **13**, 18–21 (1998).

32. Cunningham, P. & Delany, S. J. k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples). *ACM Comput. Surv.* **54**, (2020).

33. Vikramkumar, B, V. & Trilochan. Bayes and Naive Bayes Classifier. (2014).

34. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **13-17-August-2016**, 785–794 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2024.10.15.24315574; this version posted October 17, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

35. Geurts, P., Ernst, D. & Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **63**, 3–42 (2006).

36. BaggingClassifier — scikit-learn 1.5.1 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html.

37. Spearman Rank Correlation Coefficient. *Concise Encycl. Stat.* 502–505 (2008) doi:10.1007/978-0-387-32833-1_379.

38. Professor, M. S. G. Kendall's Tau. *Int. Encycl. Stat. Sci.* 713–715 (2011) doi:10.1007/978-3-642-04898-2_324.

39. Theng, D. & Bhoyar, K. K. Feature selection techniques for machine learning: a survey of more than two decades of research. *Knowl. Inf. Syst.* **66**, 1575–1637 (2024).

40. Zhao, Z., Anand, R. & Wang, M. Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform. *Proc. - 2019 IEEE Int. Conf. Data Sci. Adv. Anal. DSAA 2019* 442–452 (2019) doi:10.1109/DSAA.2019.00059.

41. Aggrawal, R. & Pal, S. Sequential Feature Selection and Machine Learning Algorithm-Based Patient's Death Events Prediction and Diagnosis in Heart Disease. *SN Comput. Sci.* **1**, 1–16 (2020).

42. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**, 389–422 (2002).

43. Dietterich, T. G. Ensemble Methods in Machine Learning. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **1857 LNCS**, 1–15 (2000).

44. Gemperli, K. *et al.* Chronic striatal cholinergic interneuron excitation induces clinically-relevant dystonic behavior in mice. *bioRxiv* 2023.07.19.549778 (2023) doi:10.1101/2023.07.19.549778.

45. ExifTool by Phil Harvey. https://www.exiftool.org/.

46. Bajpai, R. & Aravamuthan, B. SecurePose: Automated Face Blurring and Human Movement Kinematics Extraction from Videos Recorded in Clinical Settings. (2024).