

RESEARCH

Open Access



The use of machine learning and deep learning techniques to assess proprioceptive impairments of the upper limb after stroke

Delowar Hossain¹, Stephen H. Scott³, Tyler Cluff² and Sean P. Dukelow^{1*}

Abstract

Background Robots can generate rich kinematic datasets that have the potential to provide far more insight into impairments than standard clinical ordinal scales. Determining how to define the presence or absence of impairment in individuals using kinematic data, however, can be challenging. Machine learning techniques offer a potential solution to this problem. In the present manuscript we examine proprioception in stroke survivors using a robotic arm position matching task. Proprioception is impaired in 50–60% of stroke survivors and has been associated with poorer motor recovery and longer lengths of hospital stay. We present a simple cut-off score technique for individual kinematic parameters and an overall task score to determine impairment. We then compare the ability of different machine learning (ML) techniques and the above-mentioned task score to correctly classify individuals with or without stroke based on kinematic data.

Methods Participants performed an Arm Position Matching (APM) task in an exoskeleton robot. The task produced 12 kinematic parameters that quantify multiple attributes of position sense. We first quantified impairment in individual parameters and an overall task score by determining if participants with stroke fell outside of the 95% cut-off score of control (normative) values. Then, we applied five machine learning algorithms (i.e., Logistic Regression, Decision Tree, Random Forest, Random Forest with Hyperparameters Tuning, and Support Vector Machine), and a deep learning algorithm (i.e., Deep Neural Network) to classify individual participants as to whether or not they had a stroke based only on kinematic parameters using a tenfold cross-validation approach.

Results We recruited 429 participants with neuroimaging-confirmed stroke (< 35 days post-stroke) and 465 healthy controls. Depending on the APM parameter, we observed that 10.9–48.4% of stroke participants were impaired, while 44% were impaired based on their overall task score. The mean performance metrics of machine learning and deep learning models were: accuracy 82.4%, precision 85.6%, recall 76.5%, and F1 score 80.6%. All machine learning and deep learning models displayed similar classification accuracy; however, the Random Forest model had the highest numerical accuracy (83%). Our models showed higher sensitivity and specificity (AUC = 0.89) in classifying individual participants than the overall task score (AUC = 0.85) based on their performance in the APM task. We also found that variability was the most important feature in classifying performance in the APM task.

Conclusion Our ML models displayed similar classification performance. ML models were able to integrate more kinematic information and relationships between variables into decision making and displayed better classification

*Correspondence:

Sean P. Dukelow
spdukelo@ucalgary.ca

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

performance than the overall task score. ML may help to provide insight into individual kinematic features that have previously been overlooked with respect to clinical importance.

Keywords Stroke, Proprioception, Robotics, Position sense, Machine learning, Deep learning

Introduction

Proprioception is the sense of body position, motion, and force based on information from muscle spindles, Golgi tendon organs, cutaneous receptors, joint receptors, and efference copy of motor commands [1–4]. Proprioceptive impairments are common after stroke [5–7] and occur in as many as 64% of stroke survivors [8]. These impairments are associated with deficits in learning sequences of movements [9], as well as decreased independence, quality of life, and poor functional recovery [10].

Clinical assessments of proprioception have traditionally relied on coarse observer-based examinations. Most often, patients are asked to close their eyes while an examiner moves the distal part of the patient's finger, or the entire finger, up or down. The patient is asked to report the position of their fingertip/finger. Alternatively, some clinicians administer the Thumb Localization Test [11]. This is a simple test in which the clinician passively moves the patient's hand to a random position overhead while the patient's eyes are closed, and the patient must then reach to grasp their passively moved thumb with the opposite hand. Unfortunately, these clinical tests have poor reliability, lack resolution, and display ceiling effects [12, 18–20]. Some research groups have designed standardized clinician-administered tests such as the Nottingham Sensory Assessment [12], Wrist Position Sense Test (WPST) [13, 14], and Rivermead Assessment of Somatosensory Performance (RASP) [15] in attempts to deal with the issues outlined above.

However, much of the field studying proprioception has moved to the use of automated measurement tools [16, 17]. Robotic and instrumented assessments are commonly used in research studies of upper extremity proprioception [13, 21–23]. Some authors have used passive movement threshold detection paradigms [24, 25], whereas others have used single limb position-matching [26–29] or mirror-matching tasks [29–31]. Our group has significant experience using a robotic arm position matching task in individuals after stroke [31–34]. The arm position matching task, which we used in this study, relies on mirror matching and can measure various aspects of an individual's position sense, including variability in matching positions, systematic shifts in the perceived workspace, and perceived contraction or expansion of the workspace. This task can be administered quickly (~3 min) and has several advantages over typical clinical measures, including generating reliable,

continuous measures of position sense after stroke, lack of floor or ceiling effects, and the fact that the interpretation of human examiners is not required [31]. Robotic proprioceptive testing produces a rich dataset of kinematic measures that quantify impairments that can be difficult to assess at the bedside by a clinician observer.

Robotic assessments of proprioception can generate a large volume of data that may eventually be useful in predicting outcomes and planning for treatment after stroke. Machine Learning (ML) may be helpful in this regard. Several different ML methods exist, each taking advantage of different mathematical processes and algorithms. Some ML methods are more appropriate for certain types of data [35, 36]. Several past studies have attempted to predict clinical outcomes following stroke (e.g., discharge from a rehabilitation unit to home, risk of medical complications, risk of readmission to hospital) using standardized observer-based clinical scales [37–43]. Many of these studies relied on Logistic Regression, although a few used Machine Learning techniques [53–56]. ML techniques are highly effective algorithms that are driven by large volumes of data and can aid in prognosis. They are a set of powerful algorithms capable of modeling hidden and complex relationships between clinical variables and treatment outcomes without necessarily relying on any formal statistical assumptions [44]. Recently, Deep Learning-based approaches such as Deep Neural Networks, a broader family of ML techniques, have achieved impressive results across a variety of Artificial Intelligence (AI) fields [45–49]. Deep Learning approaches are inspired by the human brain's ability to abstract high-level representations from low-level sensory stimuli [50]. These multi-leveled approaches can be mathematically represented as multi-layered neural networks and recently are able to be trained in layer-wise backpropagation to obtain tractable optimization [51]. These techniques are currently state-of-the-art in applications of speech recognition, image processing, computer vision, and natural language processing [52].

ML techniques have begun to receive some attention in the field of stroke recovery. ML has been used for predictive modelling of recovery using the Barthel Index [57] and to help interpret data collected from Inertial Measurement Units in post-stroke gait [58]. There are significant opportunities for the use of ML techniques as interest continues to grow in the use of kinematic measures to quantify sensory and motor

functions. Identifying particular kinematic attributes that may not be easily recognized on clinical examination, predicting recovery following stroke and/or identifying specific kinematic attributes that might be important for targeted intervention would seem to be potential uses of ML in stroke recovery. In the current study, we sought to examine the ability of ML techniques to use kinematic data to determine whether or not an individual had a stroke. Our experimental example was chosen to provide a good test of the ability of different ML techniques to use kinematic data to classify stroke history when the outcome was known. We chose to do this with data from a proprioceptive task, as clinical assessments can be unreliable and making such a determination based on clinical data or individual kinematic parameters could prove to be substantially challenging. Further, post-stroke proprioceptive deficits can significantly impact stroke recovery. Studies have linked impaired proprioception with poorer motor recovery [59–63] and longer lengths of hospital stay [64]. Given the challenging nature of quantifying proprioception, we felt it may provide an ideal test case for the utility of ML techniques. Success at using complex kinematics from a robotic proprioceptive task to classify stroke, could bode well for future more challenging classification tasks in other neurologic conditions where the diagnosis is not yet known.

In the present study, we examined the performance of individuals with stroke and healthy controls on a robotic Arm Position Matching (APM) task. The goals of this study were: (1) to compare the different ML and DL techniques with a more traditional model that relied on the 95% cut-off score of normative data for different attributes of position sense, determining which technique flagged the highest number of stroke participants as abnormal, (2) to compare different ML and DL techniques, and their ability to classify whether someone has had a stroke or not, and (3) to examine the relative importance of different parameters measured in the APM task and their usefulness in classifying whether or not someone has had a stroke.

Methods

Participants

Participants with stroke were recruited from the inpatient acute stroke or stroke rehabilitation units at the Foothills Medical Centre, the Dr. Vernon Fanning Care Centre in Calgary, Alberta, Canada, and Providence Care, St Mary's of the Lake Hospital, Kingston, Ontario, Canada. Inclusion criteria for participants with stroke were: recent onset (<35 days) of first clinical stroke and age ≥ 18 years. Exclusion criteria for participants with stroke were: other underlying neurological conditions (e.g. Parkinson's, Multiple Sclerosis), upper limb orthopedic impairments, inability to understand task instructions or evidence of apraxia [65]. Neurologically intact control participants who also met the inclusion and exclusion criteria above, but had no history of stroke, were recruited from the communities of Calgary, Alberta, and Kingston, Ontario, Canada. This study was reviewed and approved by the University of Calgary Conjoint Health Research Ethics Board and the Queen's University Research Ethics Board. All participants gave written informed consent before performing the assessment.

Robotic assessment

Robotic Device The robotic assessment of position sense was performed using a Kinarm Exoskeleton robotic device [66] (Fig. 1A; Kinarm, Kingston, Ontario, Canada), which permits movements of the arm in the horizontal plane involving horizontal abduction/adduction of the shoulder and flexion/extension of the elbow. The planar robot has 2 degrees of freedom, articulating at both the shoulder and elbow. The device records at a rate of 1000 Hz and has a position resolution of 0.1 mm. Participants were seated in a height-adjustable wheelchair base with their arms supported against gravity. The device was fit to each participant's arm by research staff who were trained to conduct the robotic assessment. The robot was wheeled to a 2D virtual/augmented reality display. The visual display is capable of projecting virtual targets into the plane of the participant's arm during calibration and task performance. Given the focus on proprioceptive function, visual stimuli were not displayed on the screen during the experiment. Direct vision of the upper

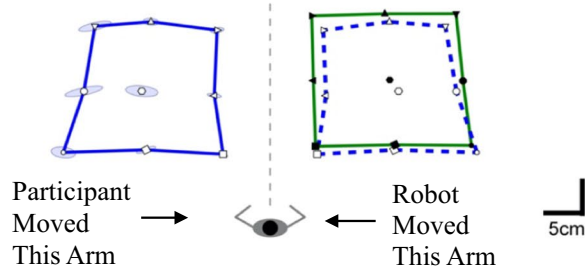
(See figure on next page.)

Fig. 1 Arm Position Matching (APM) task. **A** The Kinarm exoskeleton robot. **B** Typical healthy control participant data. The robot moved the participant's passive right hand to one of 9 spatial locations (filled symbols). The participant then attempted to mirror match with the left active hand (open symbols). The solid blue line connects the average final positions of the outer eight target locations of the matching hand (active hand). Solid green line connects the outer eight targets for the robot moved passive hand. The dashed blue line is the mirror reflection of solid blue line, which allows a visual comparison of the average final outer 8 positions of the active and passive (robot-moved) hands. Ellipses represent one standard deviation of the matched positions. The ellipses represent trial-to-trial variability, where a larger ellipse means the participant was less consistent (i.e., more variable) in matching the position of their passive hand with the active hand. **C** An exemplar stroke participant who demonstrated high variability in position matching. **D** An exemplar stroke participant who demonstrated a contracted sense of their workspace. **E** An exemplar stroke participant who demonstrated a spatial shift of their workspace

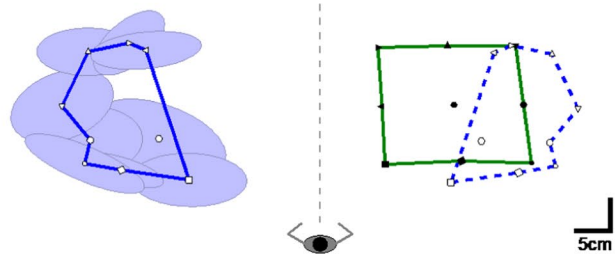
A. Kinarm Exoskeleton Robot



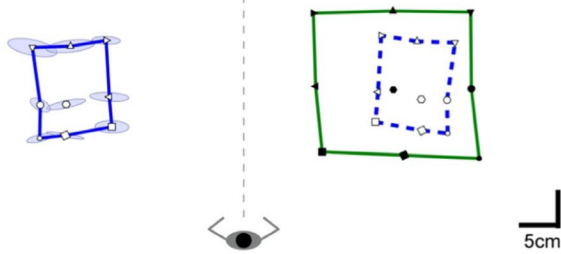
B. Healthy Control



C. Stroke – Variability



D. Stroke – Contraction/Expansion



E. Stroke - Spatial Shift

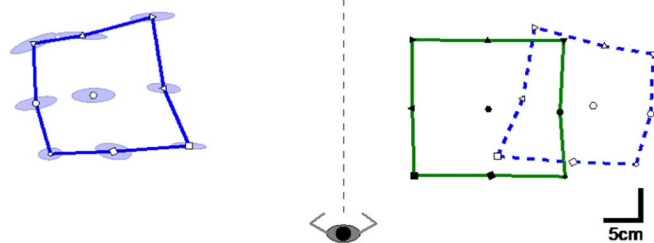


Fig. 1 (See legend on previous page.)

extremities was occluded by a shutter and a bib. The set-up and calibration procedures took between 6 and 8 min for each participant.

Arm Position Matching Task The Arm Position Matching (APM) task was used to assess the individual's position sense of their arm and has been described previously [31–34, 67]. Participants were instructed to relax one arm (passive hand) and let the robot passively move their hand to one of four/nine spatial locations separated by 20/10 cm (Fig. 1B, 9-target task). The 4-target protocol is spaced on a 2×2 grid with targets spaced at 20 cm intervals in the X- and Y-directions. The 9-target protocol is the same as the 4-target protocol but includes nine targets spaced on a 3×3 grid at 10 cm intervals. Target locations were pseudo-randomized within a block. Each block contained one trial at each target location and participants completed six blocks. The robot moved the passive hand using a bell-shaped speed profile (max speed < 1 m/s). After the robot completed the passive movement, participants were asked to move their other arm (active hand; Fig. 1B) to mirror-match the spatial position of the passive hand. Participants were granted as much time as necessary to match the active hand position with the passive hand. Participants notified the examiner when they had matched their hand position, and the examiner triggered the next trial. Each participant completed either the 4-target or 9-target task protocol [68]. For the stroke participants, the affected arm was always the passive hand. Healthy control participants completed the task twice, where each arm served as the passive hand once and we consider data from each arm as a separate participant in the analysis [69].

Robotic task parameters

The following parameters were used to quantify task performance after completing all trials: (a) trial-to-trial *Variability* (Var) of the active hand, (b) *Spatial Contraction/Expansion Ratio* (Cont/Exp) of the area matched by the active hand, (c) *Systematic Spatial Shifts* (Shift) between the passive and active hands, and (d) *Absolute Error* (AE).

Variability: *Variability* in Arm Position Matching (APM) describes the trial-to-trial consistency of the active hand in matching the same target position (Fig. 1C). It was calculated as the standard deviation of the active hand's position for each target location. The mean of the standard deviations was then calculated across all target positions in the x-coordinate (Var_x), y-coordinate (Var_y), and resultant linear variability of both coordinates (Var_{xy}):

$$Var_{xy} = \sqrt{Var_x^2 + Var_y^2} \quad (1)$$

Spatial Contraction/Expansion Ratio: *Spatial Contraction/Expansion Ratio* describes whether a participant displayed contraction or expansion of their perceived workspace (Fig. 1D). It was calculated as the matched area/range of the workspace of the active hand relative to the passive hand. This parameter was calculated for the matched x-coordinates ($Cont/Exp_x$) by finding the difference between the mean x-coordinate of the three left and three right targets for the active hand compared with the passive hand:

$$Cont/Exp_x = \frac{range_{x_{active}}}{range_{x_{passive}}} \quad (2)$$

A similar procedure was used to calculate contraction/expansion in the y-coordinate ($Cont/Exp_y$) using the range of the top and bottom three targets. Spatial contraction/expansion in both the x- and y- coordinates ($Cont/Exp_{xy}$) was calculated by finding the area spanned by the active hand for the eight border targets and then normalized by the total spatial area spanned by these same targets using the passive hand.

Systematic Spatial Shifts: *Systematic Spatial Shifts* describe constant errors between the active and passive hands (Fig. 1E). These errors were calculated as the mean error between the passive and active hands for each target position. The mean was then calculated using the means for all target locations. Systematic shifts were calculated in the x-coordinate ($Shift_x$), y-coordinate ($Shift_y$), and combined across both coordinates to provide a measure of the resultant shift in matched positions ($Shift_{xy}$):

$$Shift_{xy} = \sqrt{Shift_x^2 + Shift_y^2} \quad (3)$$

Absolute Error: *Absolute Error* describes the mean absolute distance error between the position of the active and passive hands. The mean absolute distance error between the active hand and the target position was calculated across all trials in the x-coordinate (AE_x), y-coordinate (AE_y), and combined across both coordinates (AE_{xy}):

$$AE_{xy} = \sqrt{AE_x^2 + AE_y^2} \quad (4)$$

A total of 12 parameters were used to measure performance in the arm position matching task.

Z-score: For each of the parameters above, we relied on the Dexter-E software [70] associated with the Kinarm to calculate a Z-score. The Z-score or standardized score, is the distance, measured in standard deviations, that a data point falls from the mean of the healthy cohort. Kinarm (Kinarm, Kingston, ON) [71]

uses a consistent methodology for developing normal models to calculate the Z-scores of each parameter. Parameter scores from the distribution of the normative data set (developed from neurologically intact controls) are transformed using a Box-Cox power transformation to convert the distribution to a normal distribution [72]. The transformed data are fitted by accounting for age, sex, handedness, and robotic platform (exoskeleton, endpoint robot) using Multiple Linear Regression (MLR). After the first regression, the standard deviation of the residuals is then modeled using a second MLR accounting for the same factors (age, sex, handedness, robotic platform). Z-scores are calculated using the residuals of the first regression and standard deviation modeled by second regression for each parameter. Z-scores are the particular values from the mean, i.e., a Z-score of 1 signifies that a value was 1 standard deviation above the mean, and a Z-score of -1 signifies that a value was 1 standard deviation below the mean of the healthy control data.

To ensure the distribution was “close-to-normal”, the skew and kurtosis of the final distribution were calculated and compared to the following criteria (Eqs. 5 and 6). These criteria were selected from Pearson and Please [73] so that it is statistically valid to use parametric tests with the Z-scores.

$$\text{skew} : \text{abs}(\sqrt{\beta_1}) \leq 0.8, \sqrt{\beta_1} = \frac{\mu_3}{\sigma^3} \quad (5)$$

$$\text{kurtosis} : 2.4 \leq \beta_2 \leq 3.6, \beta_2 = \frac{\mu_4}{\sigma^4} \quad (6)$$

where σ is the standard deviation, and μ_3 and μ_4 are the third and fourth moments of the mean.

Task Score: A task score gives a global measure of a participant’s performance for a given task. It measures how far the participant’s performance is from the best performance. The first stage in calculating the task score is to convert the task parameter scores into standardized Z-scores (described above). The second stage is to identify whether the best performance for a given metric reflects large negative Z-scores, large positive Z-scores, or near zero Z-scores. The Z-scores are transformed into Zeta scores using Eq. 7 for those parameters in which best performance is one-sided (i.e., large negative or large positive Z-scores).

$$\zeta = \sqrt{2} \bullet \text{erfc}^{-1}\left(\frac{1}{2} \bullet \text{erfc}\left(\frac{\pm z}{\sqrt{2}}\right)\right) \quad (7)$$

where ‘+’ is used when poor performance is positive and ‘-’ is used when poor performance is negative.

In the final stage, task scores are calculated based on the performance of healthy controls. The root-sum-square (RSS) distance of Z-scores and Zeta scores are calculated using Eq. 8 for healthy controls. RSS distance is also known as the Euclidean distance and is transformed into a Z-score using a Box-Cox transform. The Z-score of the RSS distance is then transformed to a one-sided statistic using Eq. 7.

$$\text{rssDistance} = \sqrt{\sum_i z_i^2 + \sum_j \zeta_j^2} \quad (8)$$

where $\sum_i z_i^2$ includes all two-sided parameters and $\sum_j \zeta_j^2$ includes all one-sided parameters.

Task scores are always positive. A score of 0 corresponds to the best performance, and increasing values represent poorer performance. If the task score is > 3.29 (that is normally 1 in 1000) for control participants, then that participant was classified as an outlier for the task and removed. Outliers were removed to improve the robustness of the modeling process of normative data sets.

Clinical assessments

A broad range of clinical assessments was performed to characterize the impairment of stroke participants in this study. The assessments served to quantify sensation, movement, cognition, and functional abilities. The assessments were performed by a physician or physiotherapist who had expertise in stroke rehabilitation. They were blinded to the results of the robotic assessment.

Position sense was clinically assessed using the Thumb Localization Test (TLT) [11] because it has been used to quantify whole-limb position sense in many studies involving stroke [74–82]. In this test, the examiner moves the participant’s stroke-affected arm to a position in front of the participant at or above eye level, lateral to the midline with the participant’s eyes closed. The participant is then asked to pinch the thumb of that limb with the opposite thumb and forefinger (reaching limb). Participants were scored as 0 if their performance was considered normal (completed task perfectly) to 3, which is considered markedly abnormal (the participant was unable to find his or her thumb and did not climb up the affected arm to locate it).

Motor impairment was assessed using the Purdue Peg Board test (PPB) (Lafayette Instrument Co., Lafayette, IN, USA) [83] and the Chedoke-McMaster Stroke Assessment (CMSA) [84]. In the PPB assessment, participants placed as many small pegs as possible into holes in a board over 30 s using one hand. The participant is required to use the proximal upper extremity to keep the hand in the appropriate position to retrieve and insert

each peg as a test of fine motor skills. The CMSA relies on the concept of stages of motor recovery, which was first introduced by Twitchell [85]. The CMSA classifies participants into subgroups based on the stage of motor recovery. The 7-point scale corresponds to seven stages of motor recovery, where a score of 1 is considered the most abnormal and a score of 7 is normal.

Functional abilities were assessed using the Functional Independence Measure (FIM) [86]. It is used as a metric for independence within activities of daily living. Within the 18-item scale, 13 items are considered motor tasks, and 5 items are considered cognitive tasks. In the current manuscript, we present the total FIM score (measured out of 126) and the FIM score for the motor component (measured out of 91).

Data analysis

Data analysis was done using Machine Learning and Deep Learning techniques in the Python programming language (version 3.7.4) [87]. In the first step of our analysis, we determined when stroke participants

were impaired on robotic parameters using the Z-scores described above. We determined the 95% cut-off score of control performance (Task score > 1.96 is considered as impaired and Task score ≤ 1.96 is considered as unimpaired) on each robotic parameter to find whether an individual participant failed on a given parameter. When a stroke/control participant’s score fell outside of the control range, they were classified as impaired on that robotic task. Our primary analysis compared the impairment rate found on individual parameters and the overall task score (so called cut-off score technique) versus the ability of Machine Learning and Deep Learning techniques to determine impairment. Data imputation was unnecessary as there were no missing data in the sample of control and stroke participants.

Machine learning and deep learning

Flowchart of the Classification Models: The workflow blueprint of the data classification models is shown in Fig. 2. The K-fold cross-validation (K=10, CV) training and testing data represent the outcome measures

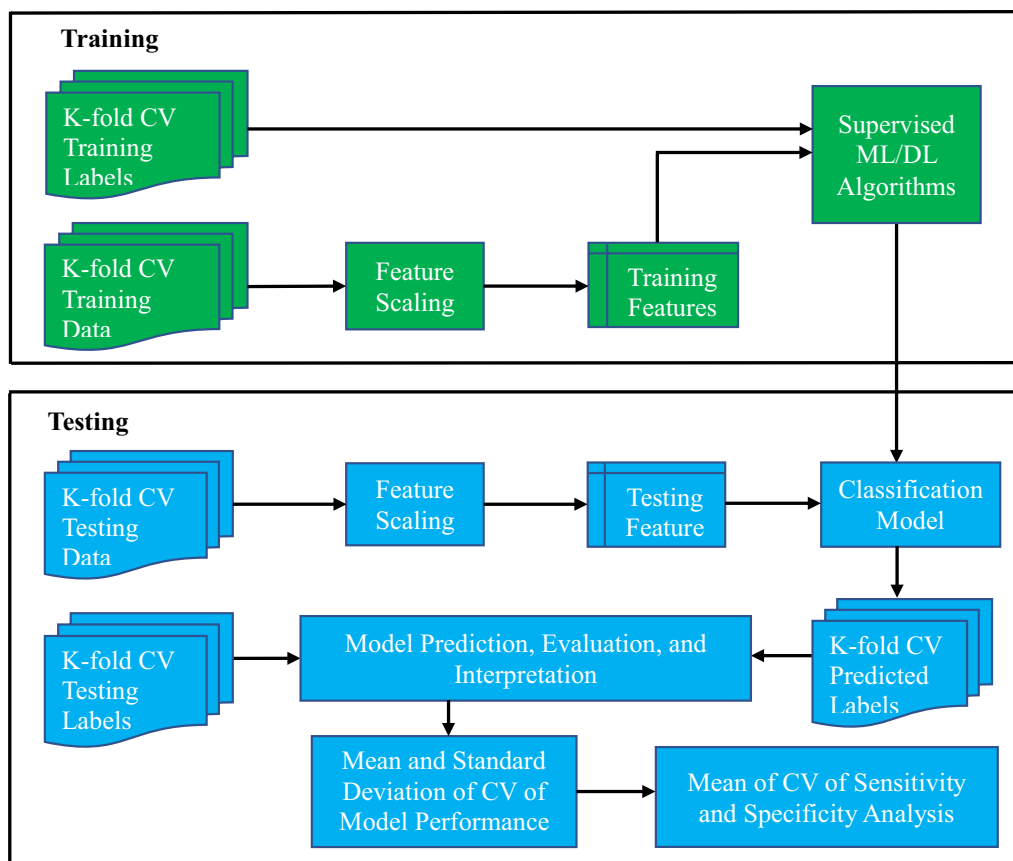


Fig. 2 The workflow of K-fold (K=10) cross-validation (CV) of the machine learning and deep learning models. The training and testing data refer to outcome measures derived from the position matching task in each stroke and control participant. The model generates a label that classified each individual participant as a control or participant with stroke

(features) derived from the Arm Position Matching (APM) task (12 parameters) of each control and stroke participant. K-fold CV training and testing data were classified and labeled into two different categories (“control” and “stroke”). This data was passed through feature extraction and scaling processes. It was then fitted to the supervised machine learning and deep learning models. After evaluation, we calculated the mean and standard deviation across the K-fold CV for all model performance metrics. Finally, we calculated receiver operating characteristic curves (ROC curves) for the mean of the K-fold cross-validated results of each model.

K-fold Cross-Validation (CV): The K-fold Cross-Validation procedure randomly divided the dataset into K-disjoint folds. One-fold was used for testing and remain K-1 folds were used for training the model. This process was repeated K-times until the testing was performed on all K folds. All folds contained equal data points unless otherwise specified. We applied K-fold cross-validation (where $K=10$) to estimate the performance and reliability of each classification algorithm and enable meaningful comparison between classification models. The performance of the classification models was evaluated by the mean and standard deviation across the K-fold datasets.

Features: A feature represents a measurable piece of data that can be used for analysis. It is also known as an “attribute” or “variable”. In our case, features were the Z-score data of the 12 task parameters (Variability X, Variability Y, Variability XY, etc.), such that all features were selected for our analysis. The features were then normalized using the min–max normalization (where the minimum value of that feature got transformed into 0, the maximum value got transformed into 1, and every other value got transformed between 0 and 1) so that the variance of the features was in the same range. Then, features were trained and tested using machine learning and deep learning models. After that, we classified participants with each model and evaluated model performance using the following metrics: accuracy, precision, recall, F1 score, receiver operating characteristics (ROC) curve, and feature importance.

Classification methods

We applied five Machine Learning (ML) techniques: Logistic Regression (LR) [88], Decision Tree (DT) [89], Random Forest (RF) [90], Random Forest with Hyperparameters Tuning (RFT) [91], and Support Vector Machine (SVM) [92]. We also applied one Deep Learning technique: Deep Neural Network (DNN) [93] for the classification (or supervised learning) of stroke and control data. We chose these 6 techniques as the type of data we were using is compatible with their application. All 6 techniques are commonly used for different ML

problems [94–97]. Further, we expected that because of the differences between the techniques, that each technique had the potential to produce a different result when analyzing the same data. We hypothesized that the Random Forest model would outperform the other models because it represents an ensemble learning method rather than a single decision tree [98, 99].

Logistic Regression (LR): We used a Logistic Regression model to classify each participant as a stroke or control based on their performance in the arm position matching task. For that purpose, we implemented a logistic regression classifier that was fitted in the binary logistic regression regularization. This regularization added a penalty as model complexity increased to ensure the model generalized the data and prevented overfitting with an increase in parameters. LR model assumes a linear relationship between the input features and output. The binary logistic model had a dependent variable with two possible outcomes as healthy control and stroke. We used a tolerance of 0.0001 and the maximum number of iterations of 100 as criteria to stop network training.

Decision Tree (DT): We implemented a Decision Tree classifier as one of predictive modeling. It uses a tree-like model in which each internal node (non-leaf) is labeled with an input feature. The arcs coming from a node (branch) labeled with an input feature are labeled with each of the possible values of the target feature or the arcs lead to a subordinate decision node on a different input feature. Each leaf node is labeled with a class either healthy control or stroke. This model splits the nodes of all available features/parameters and then selects the splits, which results in the most homogeneous sub-nodes.

Our decision tree classifier implementation consisted of the following parameters: Gini impurity as a criterion to measure the quality of split, best as a splitter to choose the best split, the maximum depth of the tree as 4, and the minimum number of samples at the leaf node as 1.

Random Forest (RF): We implemented an ensemble learning model (i.e., a Random Forest classifier). It is a classification algorithm consisting of many decision trees, which uses bagging and feature randomness when building each individual tree. It tries to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. The output of the random forest model was the class selected by most trees.

The parameters included in our implementation were: the number of estimators (the number of trees in the forest) was 100, Gini impurity as the criterion for the information gain, the minimum number of samples required to split an internal node was 2, and the minimum number of samples required to be a leaf node was 1. We examined

the different number of maximum depths of the tree for the DT model and found that the maximum depth of four gave us the best classification accuracy. The lower number made our model faster, but not as accurate, and higher number gave more accuracy but slow and risk of overfitting.

Random Forest with Hyperparameters Tuning (RFT): We tuned the hyperparameters (a hyperparameter is a parameter whose value is used to control the learning process) of the Random Forest model to determine the best hyperparameters. It relies more on experimental results than theory, and thus the best model to determine the optimal settings was by trying many different combinations to evaluate each model's performance.

The tuned hyperparameters of the random forest model were: the number of trees in the forest, the maximum number of levels in each decision tree, the maximum number of features considered for spotting a node, the minimum number of data points placed in a node before the node is split, and the minimum number of data points allowed in a leaf node.

Support Vector Machine (SVM): We implemented a Support Vector Machine (SVM) classifier. It constructed a set of hyperplanes (hyperplanes are decision boundaries that help to classify the data points) in high-dimensional space to perform the classification task. The model transformed the data to find an optimal boundary between outputs (control or stroke). A good separation is achieved by the hyperplane that had the largest distance, or functional margin, to the nearest training data point of any class.

Our implementation included the following parameters: the regularization parameter that must be strictly positive, the Radial Basis Function (RBF) type kernel, the size of the kernel cache as 200 MB, the pseudo-random number generator was used for shuffling the data for probability estimators, and tolerance of 0.001 was applied as the network stopping criterion.

Deep Neural Network (DNN): We also implemented a Deep Learning technique, namely, Deep Neural Network (DNN). It is a part of a broader family of machine learning techniques based on artificial neural networks.

Our DNN classifier implementation consisted of three hidden layers between input and output layers. The first hidden layer had 12 units with the Rectified Linear Unit (ReLU) as the activation function, the second hidden layer had 8 units with the ReLU as the activation function, and the third hidden layer had 1 unit with the sigmoid function as the activation function. We also used: binary cross-entropy as loss function, the Root Means Square propagation optimizer (RMSprop), the batch size of 10, and the number epoch of 100. We examined different numbers of epochs (i.e., 50, 100, 200, 500, 1000,

2000, 3000, 5000, 10,000, 20,000, 50,000) for the DNN model and found that 100 epochs gave us the best performance metrics. An epoch refers to the number of passes of the entire training dataset the deep learning technique has completed. The input layer had 12 units for 12 features, and the output had 1 unit to predict a 0 or 1 that maps back to the "healthy control" or "stroke" class. Each layer of nodes trained a distinct set of features based on the output of the previous layer. The feature hierarchical process of our DNN model made it capable of handling very large, and high-dimensional datasets with billions of parameters passed through nonlinear functions.

Feature importance

Feature importance [100] in machine learning refers to techniques that assign a score to each feature based on their usefulness in the classification task. The score is expressed as a percentage. We applied different feature importance techniques/calculations for the different machine learning techniques. For LR and SVM models, feature importance was based on an information-theoretic criterion, measuring the entropy in the changes of predictions, and perturbation of a given feature [101]. A perturbation, in this instance, is an analytical method to determine an approximate solution of nonlinear equations for which exact solutions cannot be obtained. For the DT, RF, and RFT models, feature importance was computed as the mean and standard deviation of the impurity decrease within each tree [102]. Impurity decrease is the total decrease in node impurity averaged over all ensemble trees and node impurity is a measure of the homogeneity of the labels at a node. In general, a higher score of feature importance means the specific feature has a large effect (importance) on the model that is being used to classify participants as "stroke" and "control", and a lower score means the specific feature has less impact on the classification model.

Results

Participant Demographics: Data were collected from 429 stroke participants and 465 healthy control participants. Demographics and clinical features of all groups are summarized in Table 1. Ninety-three percent of control and 92% of stroke participants were right-hand dominant. Two control and three stroke participants were scored as having mixed handedness on the Edinburgh Handedness Inventory [103]. Forty-eight percent of stroke participants were observed to have proprioceptive impairment based on the Thumb Localization Test (TLT) test. Seventy-six percent of stroke participants demonstrated motor impairments on their affected arm based on the Chedoke-McMaster Stroke Assessment (CMSA) test.

Table 1 Demographic and clinical information for the sample of 894 participants of healthy control and stroke

	Control (n = 465)	Stroke (n = 429)
Age	51 (20–88)	63 (18–92)
Sex	244 M, 221 F	280 M, 149 F
Dominant Hand	434 R, 29 L, 2 A	393 R, 33 L, 3 A
Days since Stroke	...	17 (1–34)
Types of Stroke	...	370 I, 59 H
TLT [0, 1, 2, 3]		
Affected Side	...	[210, 104, 73, 30]*
CMSA [1–7]		
Affected Arm	...	[29, 33, 51, 58, 61, 80, 103]
Unaffected Arm	...	[0, 0, 0, 0, 15, 94, 320]
CMSA [1–7]		
Affected Hand	...	[31, 33, 37, 43, 74, 93, 104]
Unaffected Hand	...	[0, 0, 0, 0, 7, 125, 297]
PPB		
Affected Side	...	6.9 (0–17.5)
Unaffected Side	...	10.5 (2.5–19)
FIM (Total Score)	...	93.7 (37–126)
FIM (Motor)	...	65 (13–91)

Data are presented as the mean (range) unless otherwise noted. Square brackets for TLT, CMSA scores indicate the actual number of individuals who obtained a given score on the test, e.g., 210 individuals scored 0 on the TLT

M—Male; F—Female; R—Right; L—Left; A—Ambidextrous; H—Hemorrhagic; I—Ischemic; TLT—Thumb Localizing Test; CSMA—Chedoke-McMaster Stroke Assessment; PPB—Purdue Peg Board; and FIM—Functional Independence Measure

*2 scores were missing

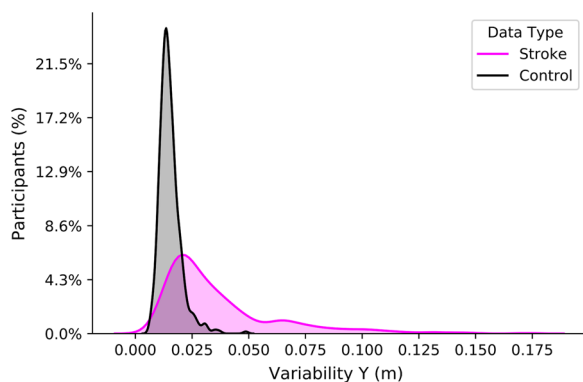


Fig. 3 Histogram showing the distribution of *Variability Y* values in healthy controls and participants with stroke. The percentage on the y-axis is the participant count in each bin normalized to the number of participants with stroke (n = 429) and control (n = 465). An Additional file 1: Supplementary figure 1 is available and demonstrates the histogram distribution of all parameters of healthy controls and stroke participants

Data Visualization: To visualize the distribution of scores on the robotic task parameters, we plotted

histograms of each parameter. *Variability Y* for stroke and control participants is presented in Fig. 3. This exemplar figure demonstrated that the distribution of values of the *Variability Y* parameter overlapped between stroke and control participants. We chose to present *Variability Y* because this parameter had the most influence on the classification tasks (see Fig. 6). Similar findings were seen when examining the distributions of the other parameters (see Additional file 1: Figure S1). The overlap of stroke and control data highlighted the challenge of differentiating normal from abnormal behavior based on a single parameter.

Cut-off Score Technique, Machine Learning, and Deep Learning Classifier Models: We first examined the impairment rates for individual robotic parameters and the overall task score using the 95% cut-off score technique. Figure 4A shows the mean and standard deviation of impairment rates from the tenfold cross-validation of individual parameters (*Variability*, *Contraction/Expansion ratio*, *Shift*, and *Absolute Error*) and the overall task score to find the number of impaired participants. The result indicates that the highest number of participants were impaired on the parameter *Variability XY* (48.4%) followed by the other two variabilities: *Variability Y* (47.1%) and *Variability X* (45.2%) parameters using the cut-off technique based on individual parameter. The least number of participants were impaired on parameter *Shift X* (10.9%), followed by *Shift Y* (11.5%) and *Shift XY* (20.3%). The overall task score impairment rate was 44% based on the 95% cut-off score technique.

We then went on to examine the mean and standard deviation of tenfold cross-validated performance metrics i.e., accuracy, precision, recall, and F1 score [104] of the Machine Learning and Deep Learning models (Fig. 5B). The results of this analysis indicated that RF and RFT had the highest and nearly similar accuracy (83.0% and 82.8%). In the case of precision, LR had a higher value of 86.6% than any other classifier. Again, for recall (78.0%) and F1 score (81.4%) metrics, RF had a higher value of than any other classifier. In terms of standard deviation, LR had the highest spread out over a range of 4.1% for accuracy, 7.8% for recall, and 4.9% for F1 score, whereas DT had the highest spread out over a range of 5.6% for the precision compared with other classifiers.

Receiver operating characteristic (ROC) curve: We implemented five Machine Learning classifier models, namely Logistic Regression (LR), Decision Tree (DT), Random Forest (RT), Random Forest with Hyperparameters Tuning (RFT), Support Vector Machine (SVM), and one Deep Learning classifier model, namely Deep Neural Network (DNN) to classify data into two categories: “control” and “stroke”. The mean ROC curve and Area Under the Curve (AUC) value for the classification

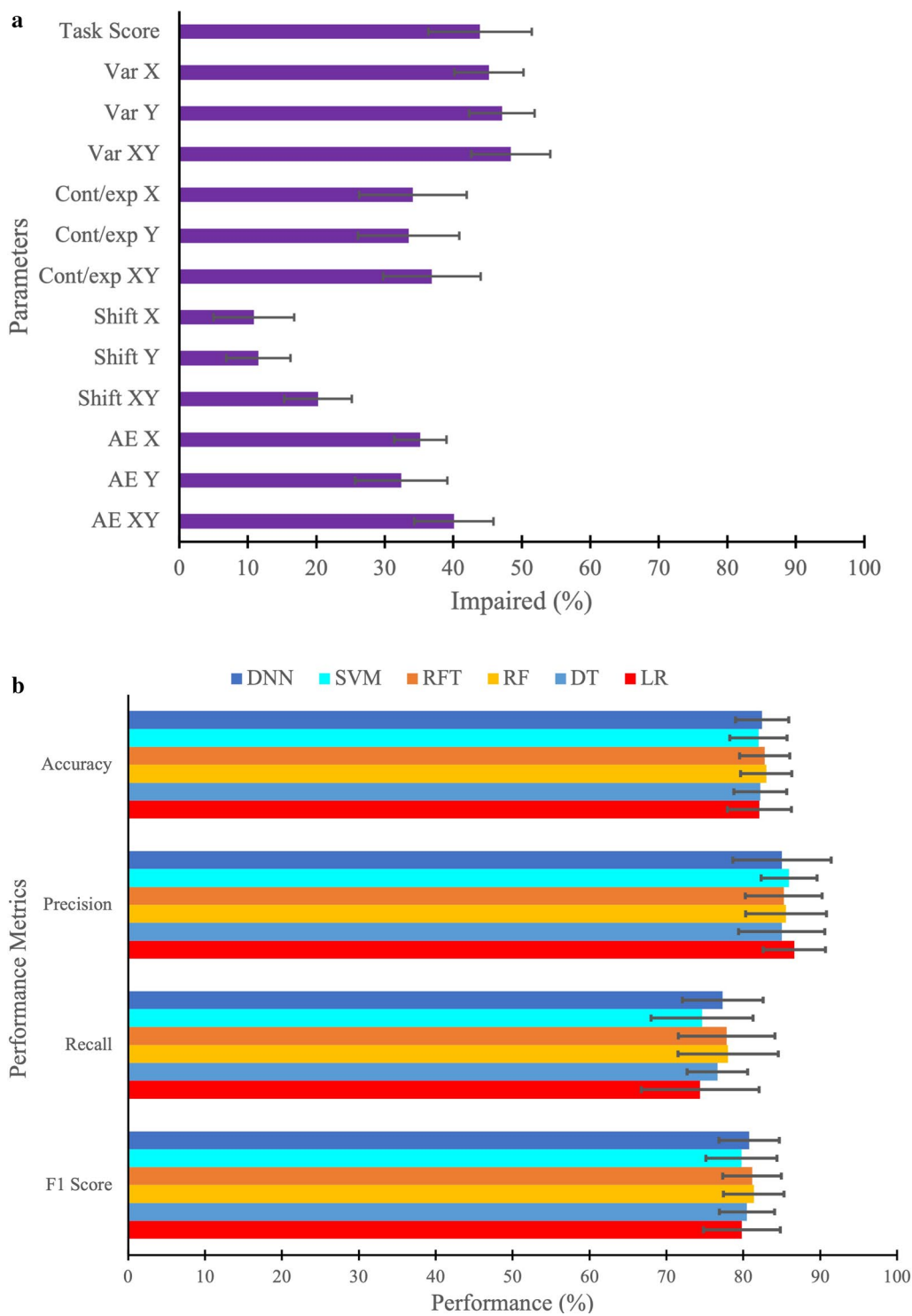


Fig. 4 **A** Performance metrics when classified based on individual parameters (Var: variability, Cont/Exp: contraction/expansion, Shift, AE: absolute error) of arm position matching task, as well as overall task score to find the number of impaired participants. **B** Performance metrics (accuracy, precision, recall, and F1 score) for the machine learning and deep learning models (LR: Logistic Regression, DT: Decision Tree, RF: Random Forest, RFT: Random Forest with Hyperparameters Tuning, SVM: Support Vector Machine, DNN: Deep Neural Network)

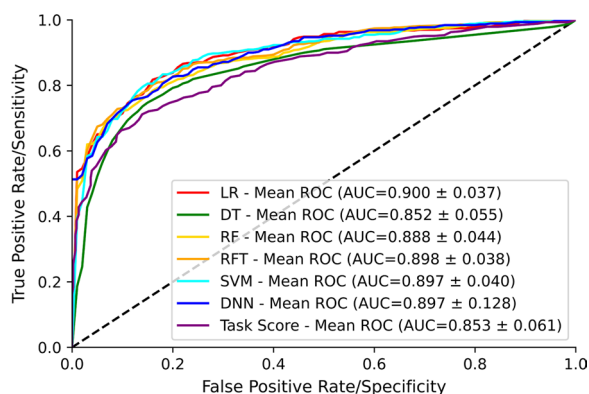


Fig. 5 Mean of tenfold cross-validation of the receiver operating characteristic (ROC) curve and area under the curve (AUC) for the classification performance of LR, DT, RF, RFT, SVM, DNN, and task score. The dashed line corresponds to classification due to random chance (AUC = 0.5, i.e., 50% sensitivity and 50% specificity)

performance of LR, DT, RF, RFT, SVM, and DNN models is shown in Fig. 5A. We also present the ROC and AUC for the overall task score which is calculated from individual parameter scores as outlined in the methods for comparison. The best possible classifier model would yield a point in the upper-left coordinate (0,1) of the ROC space, representing 100% sensitivity and 100% specificity, which is called a perfect classification. The LR model had the highest AUC (AUC = 0.900) value for the classification task, suggesting that LR had the best separation capability between control and stroke, followed by the RFT model (AUC = 0.898). RFT model performed slightly better at classification than RF (AUC = 0.888). SVM and DNN models performed similarly to the RFT model (AUC = 0.897). The DT performed the worst at the classification task among all models (AUC = 0.852). In summary, LR had the highest level of sensitivity and specificity, whereas DT had the lowest level of specificity and sensitivity among models. For comparison, the overall task score had an AUC = 0.853.

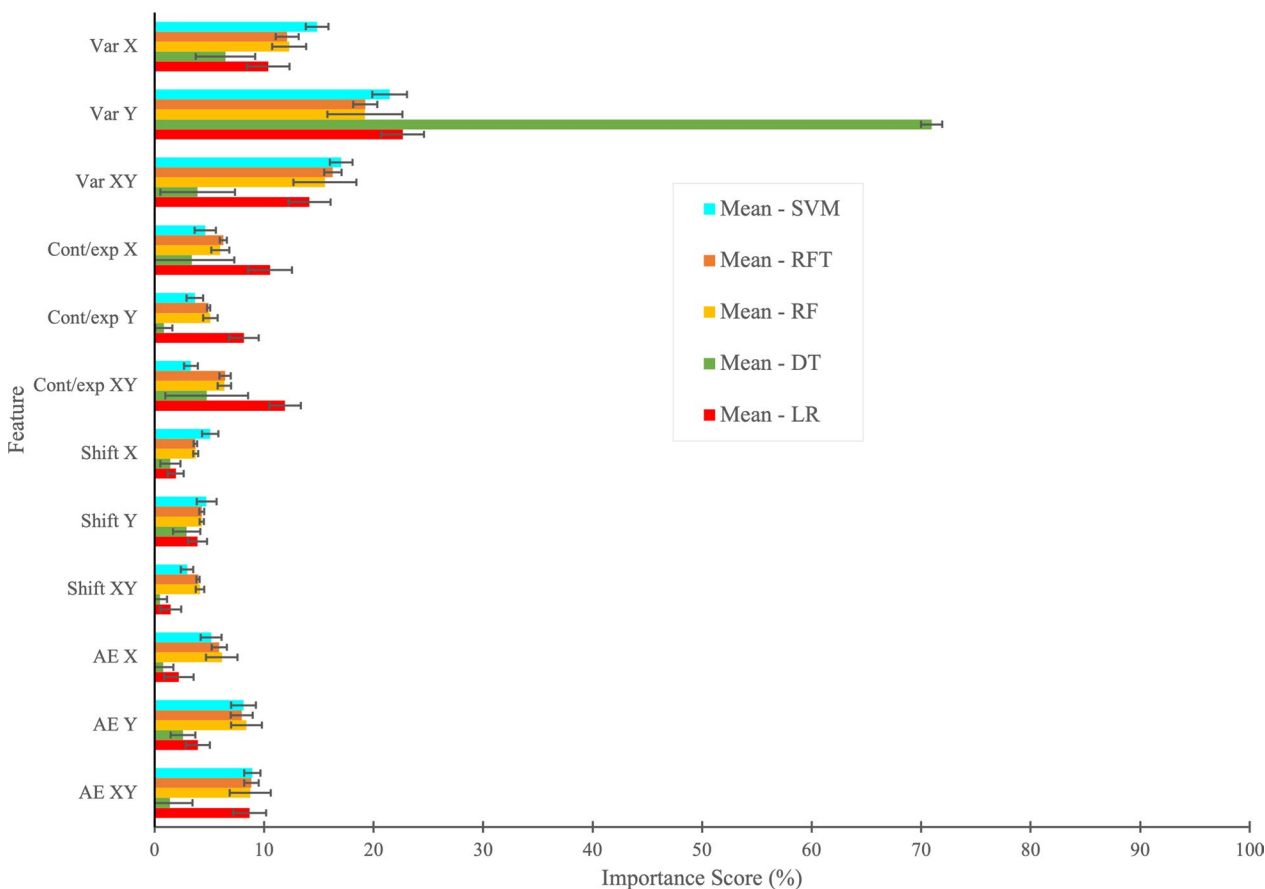


Fig. 6 Mean and standard deviation of K-fold (K = 10) cross-validation of feature importance based on individual parameters (Var, Cont/exp, Shift, and AE) obtained from LR, DT, RF, RFT, and SVM models. Due to the complex structure of Deep Neural Network (DNN) model, we could not plot the feature importance using DNN model

Feature Importance: The mean and standard deviation of tenfold cross-validation of feature importance based on individual parameters (Variability, Contraction/Expansion ratio, Shift, Absolute Error) obtained using LR, DT, RF, RFT, and SVM models for the classification task is shown in Fig. 6. We were not able to plot feature importance using the DNN model because of its complex structure. We can see that different models had different feature importance scores in percentage. Across the models, some features tended to have higher feature importance scores, whereas others tended to have lower feature importance scores. For instance, we observed that *Variability Y* was the most important feature of all models for the classification task, followed by *Variability XY* and *Variability X*. The least important feature tended to be *Shift XY*, followed by *Shift X* and *Shift Y*. Although, many features contributed similarly for the classification task using LR, RF, RFT, and SVM models, the relative importance of the features appeared to be different in the DT model (see Fig. 6).

Discussion

Proprioceptive impairment is a common consequence of stroke. Traditional clinical approaches to assess proprioception have known issues with reliability [18] and tend to rely on simplistic observer-based ordinal scales. This has led to the development of different instrumented assessment tools [31] [13, 105]. Assessments such as robotics, which can provide detailed kinematic measures, have the potential to offer new insights into the nature and severity of the proprioceptive impairments that occur after stroke. Employing machine learning techniques to explore the complex datasets that are generated by robotic assessments may prove valuable in this regard. ML techniques may help to highlight features in the data that may be important, but not obvious with other types of analyses.

The current manuscript represents our first foray into using ML techniques to classify robotic proprioceptive kinematic data from stroke survivors. We attempted to contrast the prevalence of proprioceptive deficits as identified by a previously used [31] standardized technique that relies on cut-off scores based on the 95% distribution of healthy control performance to the information derived from several different machine learning techniques. Using the cut-off score technique, we observed that the percentage of individuals classified as impaired on any given task parameter was between 10.9% and 48.4%. Variability parameters tended to detect the largest number of impaired individuals. Neuroimaging studies have shown performance on these parameters correspond with several areas of the brain including S1, posterior parietal cortex, supramarginal gyrus, superior

temporal gyrus, transverse temporal gyrus, and arcuate fasciculus [106, 107]. Using an overall task score cut-off developed from the individual parameters which were equally weighted, 44% of individuals were deemed impaired. In order to more directly compare the cut-off score performance to what can be derived from ML techniques, we examined the area under the ROC curves using the overall task score and determined that accuracy of the overall task score was 85%. The machine learning techniques, on the other hand, demonstrated comparatively higher sensitivity and specificity, with areas under the ROC curves ranging from ~85% to 90.0% (depending on the given technique) when trying to classify whether participants did or did not have a stroke based on their performance in the robot. Machine learning, by taking advantage of different parameter weightings and the relationships between parameters, performs better at classifying whether someone had a stroke. The cut-off score technique, on the other hand, tries to determine who does or does not have a proprioceptive impairment based on robotic performance (i.e., prevalence of proprioceptive deficits).

The cut-off score technique allows examination of individual kinematic variables in comparison to healthy controls (e.g., Variability_{xy}), whereas the machine learning techniques, as employed, do not. In itself, individual parameters may be important for appreciating the nature of a patient's deficits at the granular level required to precisely design an intervention. However, working with clinicians and researchers over the years, our group has been asked repeatedly to develop an overall task score that can be quickly and easily interpreted and potentially used as a primary outcome for clinical trials. The overall task score that was developed relies on summing the individual components, assuming equal weighting for each parameter, to determine a single score [108]. Despite the mathematical complexity in generating normative scores, this method is simplistic in its implementation as all parameters are equally weighted which may or may not be the most appropriate method to generate an overall score, particularly as some parameters may be highly correlated.

Machine learning techniques, as we employed them, did not calculate the prevalence of proprioceptive deficits like our cut-off score technique. Rather, we calculate and compare the accuracy, precision, recall and F1 score of the different ML techniques in attempting to classify whether someone had a stroke or not based on participant performance in the behavioural task. In general, all of the machine learning techniques we employed performed reasonably well. Perhaps this is not surprising as the machine learning techniques developed weighting values for the individual parameters (see Fig. 6), unlike

the equal weighting assigned when generating overall task scores. When exposed to new data, machine learning models learn, grow, modify, change, and develop by themselves. Simply put, machine learning and deep learning techniques do this process by leveraging algorithms that learn from data in an iterative process, which is not possible using traditional data analysis methods. These differences considered, the ML models performed only slightly better than the overall task score.

Determining whether someone has had a stroke or not based on their performance in a robotic task, on the surface, may seem a bit impractical as the majority of cases of stroke already have a diagnosis based on clinical observations and confirmed with some form of neuroimaging (either computed tomography or magnetic resonance imaging). In the current manuscript, however, knowledge of the diagnosis provided a ground truth for us to test the performance of machine learning and make comparisons to the way which we have historically analyzed robotic data. While we do not see using machine learning to make the diagnosis of stroke using kinematics to be practical, our study demonstrates the potential utility of machine learning tools. Going forward, we foresee the utility of machine learning in being able to predict recovery patterns. There is also potential utility in classifying other disease conditions based on kinematics where the diagnosis is not always as clear cut as stroke (e.g., Movement Disorders) for which we have neuroimaging evidence.

In the present manuscript, the machine learning techniques we employed, for the most part, performed similarly. While these similarities were perhaps not so surprising based on other published studies that have shown relatively low variability in the results produced by different machine learning techniques [109–113], there are differences between how some of these techniques were mathematically operationalized (e.g., Logistic Regression vs. Support Vector Machines). These differences ultimately led to variability in the features/parameters the techniques considered most important (Fig. 6). In retrospect, our group would have been challenged to predict, a priori, why some features were deemed more important than others from what we have seen of previous studies based on neuroanatomic or neurophysiologic underpinnings. However, the highest variability and impairments rates in our dataset were noted to be in the parameter Variability Y which, in hindsight, explains why this feature was so important to many of the ML techniques. While overfitting can be a concern when using machine learning, we used a relatively large dataset with a cross-validation approach to minimize the risk of this. In the end, the similar performance of the different

models may simply stem from the fact that the underlying dataset used was the same.

Our study is not without its limitations. The APM task we employed requires both hemispheres and as such may be more likely to be impacted by a stroke lesion, than other tasks that can be used to study proprioception such as the just noticeable difference method. Individuals with hemispatial neglect may have difficulty performing the arm position matching task [114] and the results of APM in an individual with hemispatial neglect can be difficult to disentangle from those of someone with isolated proprioceptive loss as the two impairments often co-exist. In the present study, we did not provide the ML algorithms with information about the presence or absence of hemispatial neglect as this would have ensured correct classification on the basis of the presence of the syndrome. Further, the ML techniques we employed required hundreds of datasets from individual participants which required thousands of hours to collect. In general, these techniques do not respond well to missing or unknown data, although we did not test this in the present study as we had no missing data. Like most analysis techniques used, there is always some chance of error inherent in the predictions that are made. If there is bias in the data that is used to train the machine learning techniques, this can be carried over to when a model is deployed for testing. In the present paper, we employed tenfold cross-validated datasets as this has been recommended to achieve stable results compared to traditional statistical analyses (e.g. T-test, ANOVA, Chi-Square) [115]. We then tested the model on a dataset that was not used in training to decrease the risk of bias.

As neurorehabilitation begins to incorporate more technology in both clinical practice and research settings, the opportunities for employing techniques like machine learning will continue to grow and evolve. Thoughtful application of these techniques may provide new insights into longstanding clinical problems. We could potentially see these types of techniques being used to identify prognostic factors for recovery following stroke or other disease states. There are already examples of this elsewhere in healthcare: clinical variables to predict post-stroke outcomes [113, 116–120], medical imaging diagnosis [121–124], drug discovery and manufacturing [125–127], identifying diseases and diagnosis [128, 129], outbreak prediction [130, 131], etc. Further, artificial intelligence may be helpful in guiding rehabilitation for individual patients based on their characteristics using information gleaned from thousands of patients' journeys. Machine learning provides us with a newer set of analysis tools to enhance our understanding of data. Careful implementation could lead to significant changes in the way we carry

out stroke rehabilitation in the future, but more study is needed.

Conclusions

In this work, we applied five machine learning techniques (i.e., LR, DT, RF, RFT, and SVM) and one deep learning technique (i.e., DNN) to classify stroke patients from control participants using kinematic information from a robotic assessment of proprioception. The resulting AUC of the ROC curve can range up to 90% depending on the classifiers used. Also, we were able to find the important features, which contributed significantly to the classification task. The machine learning and deep learning models we demonstrate here can be readily applied to other clinical and medical research. We see them potentially being useful for classification using kinematics in situations where the diagnosis is, as yet, unknown and for identifying data elements (features) that are not immediately obvious as important in more classical analysis techniques. Future studies may allow the prediction of recovery using ML models.

Abbreviations

AE	Absolute error
APM	Arm position matching
AUC	Area under the curve
AI	Artificial intelligence
CMSA	Chedoke-McMaster stroke assessment
Cont/Exp	Contraction/expansion
CV	Cross-validation
DL	Deep learning
DNN	Deep neural network
DT	Decision tree
FIM	Functional independence measure
LR	Logistic regression
ML	Machine learning
MLR	Multiple linear regression
PPB	Purdue peg board
RASP	Rivermead assessment of somatosensory performance
ReLU	Rectified linear unit
RMSprop	Root means square propagation optimizer
RF	Random forest
RFT	Random forest with hyperparameters tuning
ROC	Receiver operating characteristics
RSS	Root-sum-square
SD	Standard deviation
SVM	Support vector machine
TLT	Thumb localization test
Var	Variability
WSPT	Wrist position sense test

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12984-023-01140-9>.

Additional file 1: Supplementary Figure 1: Histogram showing the distribution of all parameters in healthy controls and participants with stroke. The percentage on the y-axis is the participant count in each bin normalized to the number of participants with stroke ($n = 429$) and control ($n = 465$).

Acknowledgements

We would like to acknowledge the efforts and support from M Piitz and H Bretzke in data collection and technical assistance.

Author contributions

DH analyzed the collected data, carried out the machine learning and deep learning analysis, drafted the manuscript. SHS participated in the task design, data analysis, and drafting manuscript. TC participated in data analysis and drafting of the manuscript. SPD was involved in the task design and patient recruitment, participated in data analysis, and drafted the manuscript. All authors read the final manuscript and approved it.

Funding

This research was supported by CIHR in Stroke Rehabilitation Research (MOP 106662), and a Heart and Stroke Foundation of Canada Grant-in-Aid (G-13-0003029) to SPD and an Ontario Research Foundation—Research Excellence grant (ORE-RE 04-47) to SHS.

Availability of data and materials

The dataset generated for the current study are not publicly available. Data may be available from the corresponding author on reasonable request.

Declarations

Competing interests

SHS is the co-founder and Chief Scientific Officer of Kinarm (formally known as BKIN Technologies), the company that commercializes the Kinarm robotic device used in this study. All other authors confirm no conflict of interest.

Ethics approval and consent to participate

The protocol was approved by the University of Calgary Conjoint Health Research Ethics Board and the Queen's University Research Ethics Board. All participants gave written informed consent before performing the assessment.

Consent for publication

Consent to publish was obtained from all participants; and also, consent was obtained for Fig. 1A.

Competing of interests

SHS is the co-founder and Chief Scientific Officer of Kinarm (formally known as BKIN Technologies), the company that commercializes the Kinarm robotic device used in this study. All other authors confirm no conflict of interest.

Author details

¹Department of Clinical Neuroscience, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. ²Faculty of Kinesiology, University of Calgary, Calgary, AB, Canada. ³Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada.

Received: 1 April 2022 Accepted: 18 January 2023

Published online: 27 January 2023

References

- Campbell WW, DeJong RN. *DeJong's the neurologic examination* (No. 2005). Lippincott Williams & Wilkins. 2005.
- Lephart SM. Introduction to the sensorimotor system. *Proprioception and neuromuscular control in joint stability*, 2000; pp.16–26.
- Riemann BL, Lephart SM. The sensorimotor system, part I: the physiologic basis of functional joint stability. *J Athl Train*. 2002;37(1):71.
- Riemann BL, Lephart SM. The sensorimotor system, part II: the role of proprioception in motor control and functional joint stability. *J Athl Train*. 2002;37(1):80.
- Kusoffsky A, Wadell I, Nilsson BY. The relationship between sensory impairment and motor recovery in patients with hemiplegia. *Scand J Rehabil Med*. 1982;14(1):27–32.

6. Campfens SF, Zandvliet SB, Meskers CG, Schouten AC, van Putten MJ, van der Kooij H. Poor motor function is associated with reduced sensory processing after stroke. *Exp Brain Res*. 2015;233(4):1339–49.
7. Carey LM. Somatosensory loss after stroke. *Crit Rev Phys Rehabil Med*. 1995;7(1):51.
8. Connell LA, Lincoln NB, Radford KA. Somatosensory impairment after stroke: frequency of different deficits and their recovery. *Clin Rehabil*. 2008;22(8):758–67.
9. Russo C, Veronelli L, Casati C, Monti A, Perucca L, Ferraro F, Corbo M, Vallar G, Bolognini N. Explicit motor sequence learning after stroke: a neuropsychological study. *Exp Brain Res*. 2021;239(7):2303–16.
10. Carey LM, Matyas TA. Frequency of discriminative sensory loss in the hand after stroke in a rehabilitation setting. *J Rehabil Med*. 2011;43(3):257–63.
11. Hirayama K, Fukutake T, Kawamura M. 'Thumb localizing test' for detecting a lesion in the posterior column–medial lemniscal system. *J Neurol Sci*. 1999;167(1):45–9.
12. Lincoln NB, Jackson JM, Adams SA. Reliability and revision of the Nottingham Sensory Assessment for stroke patients. *Physiotherapy*. 1998;84(8):358–65.
13. Carey LM, Oke LE, Matyas TA. Impaired limb position sense after stroke: a quantitative test for clinical use. *Arch Phys Med Rehabil*. 1996;77(12):1271–8.
14. Carey LM, Matyas TA, Oke LE. Evaluation of impaired fingertip texture discrimination and wrist position sense in patients affected by stroke: comparison of clinical and new quantitative measures. *J Hand Ther*. 2002;15(1):71–82.
15. Winward CE, Halligan PW, Wade DT. The Rivermead Assessment of Somatosensory Performance (RASP): standardization and reliability data. *Clin Rehabil*. 2002;16(5):523–33.
16. Squeri V, Zenzeri J, Morasso P, Basteris A. Integrating proprioceptive assessment with proprioceptive training of stroke patients. In *2011 IEEE International Conference on Rehabilitation Robotics* 2011; 1–6. IEEE.
17. Niessen MH, Veeger DH, Meskers CG, Koppe PA, Konijnenbelt MH, Janssen TW. Relationship among shoulder proprioception, kinematics, and pain after stroke. *Arch Phys Med Rehabil*. 2009;90(9):1557–64.
18. Lincoln NB, Crow JL, Jackson JM, Waters GR, Adams SA, Hodgson P. The unreliability of sensory assessments. *Clin Rehabil*. 1991;5(4):273–82.
19. Dellon AL, Mackinnon SE, Crosby PM. Reliability of two-point discrimination measurements. *J Hand Surg*. 1987;12(5):693–6.
20. Rinderknecht MD, Lamberg O, Raible V, Büsching I, Sehle A, Liepert J, Gassert R. Reliability, validity, and clinical feasibility of a rapid and objective assessment of post-stroke deficits in hand proprioception. *J Neuroeng Rehabil*. 2018;15(1):1–15.
21. Goble DJ, Brown SH. Dynamic proprioceptive target matching behavior in the upper limb: effects of speed, task difficulty and arm/hemisphere asymmetries. *Behav Brain Res*. 2009;200(1):7–14.
22. Scheidt RA, Lillis KP, Emerson SJ. Visual, motor and attentional influences on proprioceptive contributions to perception of hand path rectilinearity during reaching. *Exp Brain Res*. 2010;204(2):239–54.
23. Goble DJ, Coxon JP, Van Impe A, Geurts M, Van Hecke W, Sunaert S, Wenderoth N, Swinnen SP. The neural basis of central proprioceptive processing in older versus younger adults: an important sensory role for right putamen. *Hum Brain Mapp*. 2012;33(4):895–908.
24. Bengtson MC, Mrotek LA, Stoeckmann T, Ghez C, Scheidt RA. The arm motion detection (AMD) test. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2014; 5349–5352. IEEE.
25. Simo LS, Ghez C, Botzer L, Scheidt RA. A quantitative and standardized robotic method for the evaluation of arm proprioception after stroke. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2011; 8227–8230. IEEE.
26. Deblock-Bellamy A, Batcho CS, Mercier C, Blanchette AK. Quantification of upper limb position sense using an exoskeleton and a virtual reality display. *J Neuroeng Rehabil*. 2018;15(1):1–11.
27. Lowrey CR, Blazevski B, Marnet JL, Bretzke H, Dukelow SP, Scott SH. Robotic tests for position sense and movement discrimination in the upper limb reveal that they each are highly reproducible but not correlated in healthy individuals. *J Neuroeng Rehabil*. 2020;17(1):1–13.
28. Gurari N, Drogos JM, Dewald JP. Individuals with chronic hemiparetic stroke can correctly match forearm positions within a single arm. *Clin Neurophysiol*. 2017;128(1):18–30.
29. Goble DJ. Proprioceptive acuity assessment via joint position matching: from basic science to general practice. *Phys Ther*. 2010;90(8):1176–84.
30. Smorenburg AR, Ledebt A, Deconinck FJ, Savelsbergh GJ. Practicing a matching movement with a mirror in individuals with spastic hemiplegia. *Res Dev Disabil*. 2013;34(9):2507–13.
31. Dukelow SP, Herter TM, Moore KD, Demers MJ, Glasgow JJ, Bagg SD, Norman KE, Scott SH. Quantitative assessment of limb position sense following stroke. *Neurorehabil Neural Repair*. 2010;24(2):178–87.
32. Dukelow SP, Herter TM, Bagg SD, Scott SH. The independence of deficits in position sense and visually guided reaching following stroke. *J Neuroeng Rehabil*. 2012;9(1):1–13.
33. Kenzie JM, Semrau JA, Hill MD, Scott SH, Dukelow SP. A composite robotic-based measure of upper limb proprioception. *J Neuroeng Rehabil*. 2017;14(1):1–12.
34. Findlater SE, Hawe RL, Semrau JA, Kenzie JM, Amy YY, Scott SH, Dukelow SP. Lesion locations associated with persistent proprioceptive impairment in the upper limbs after stroke. *NeuroImage Clin*. 2018;20:955–71.
35. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. 2022;23(1):40–55.
36. Ray S. A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)* 2019; 35–39. IEEE.
37. Roth EJ, Lovell L, Harvey RL, Heinemann AW, Semik P, Diaz S. Incidence of and risk factors for medical complications during stroke rehabilitation. *Stroke*. 2001;32(2):523–9.
38. Ottenbacher KJ, Smith PM, Illig SB, Fiedler RC, Gonzales V, Granger CV. Characteristics of persons rehospitalized after stroke rehabilitation. *Arch Phys Med Rehabil*. 2001;82(10):1367–74.
39. Wilson DB, Houle DM, Keith RA. Stroke rehabilitation: a model predicting return home. *West J Med*. 1991;154(5):587.
40. Paolucci S, Bragoni M, Coiro P, De Angelis D, Fusco FR, Morelli D, Venturiero V, Pratesi L. Is sex a prognostic factor in stroke rehabilitation? A matched comparison. *Stroke*. 2006;37(12):2989–94.
41. Berlowitz DR, Hoenig H, Cowper DC, Duncan PW, Vogel WB. Impact of comorbidities on stroke rehabilitation outcomes: does the method matter? *Arch Phys Med Rehabil*. 2008;89(10):1903–6.
42. Kristensen HK, Tistad M, Koch LV, Ytterberg C. The importance of patient involvement in stroke rehabilitation. *PLoS ONE*. 2016;11(6): e0157149.
43. Picena MC, Recupero E, Finocchiaro F, Santagati A, Greco S, Longo P, Manca M, Cosentino E, Mayer F, Biondi T, Mugelli C. Outcome predictors of rehabilitation for first stroke in the elderly. *Eur J Phys Rehabil Med*. 2008;44:3–11.
44. Bishop CM. *Neural networks for pattern recognition*. Oxford University Press; 1995.
45. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25:1097–105.
46. Ciregan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. In *2012 IEEE conference on computer vision and pattern recognition* 2012; 3642–3649. IEEE.
47. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. 2017;19:221–48.
48. Dong D, Wu H, He W, Yu D, Wang H. Multi-task learning for multiple language translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* 2015; pp. 1723–1732.
49. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. 2015. *arXiv preprint arXiv:1511.06434*.
50. Zador AM. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat Commun*. 2019;10(1):1–7.
51. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44.
52. Du H, Ghassemi MM, Feng M. The effects of deep network topology on mortality prediction. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 2016; 2602–2605. IEEE.

53. Yang G, Deng J, Pang G, Zhang H, Li J, Deng B, Pang Z, Xu J, Jiang M, Liljeberg P, Xie H. An IoT-enabled stroke rehabilitation system based on smart wearable armband and machine learning. *IEEE J Transl Eng Health Med*. 2018;6:1–10.
54. Mohanty R, Sinha AM, Remsik AB, Dodd KC, Young BM, Jacobson T, McMillan M, Thoma J, Advani H, Nair VA, Kang TJ. Machine learning classification to identify the stage of brain-computer interface therapy for stroke rehabilitation using functional connectivity. *Front Neurosci*. 2018;12:353.
55. Lee MH, Siewiorek DP, Smailagic A, Bernardino A. Opportunities of a machine learning-based decision support system for stroke rehabilitation assessment. 2020. *arXiv preprint arXiv:2002.12261*.
56. Miao S, Shen C, Feng X, Zhu Q, Shorfuazzaman M, Lv Z. Upper limb rehabilitation system for stroke survivors based on multi-modal sensors and machine learning. *IEEE Access*. 2021;9:30283–91.
57. Campagnini S, Liuzzi P, Mannini A, Basagni B, Macchi C, Carrozza MC, Cecchi F. Cross-validation of predictive models for functional recovery after post-stroke rehabilitation. *J Neuroeng Rehabil*. 2022;19(1):1–11.
58. Wang FC, Chen SF, Lin CH, Shih CJ, Lin AC, Yuan W, Li YC, Kuo TY. Detection and classification of stroke gaits by deep neural networks employing inertial measurement units. *Sensors*. 2021;21(5):1864.
59. Broeks JG, Lankhorst GJ, Rumping K, Prevo AJH. The long-term outcome of arm function after stroke: results of a follow-up study. *Disabil Rehabil*. 1999;21(8):357–64.
60. De Weerd W, Lincoln NB, Harrison MA. Prediction of arm and hand function recovery in stroke patients. *Int J Rehabil Res*. 1987;10:110–2.
61. Morris JH, Van Wijck F, Joice S, Donaghy M. Predicting health related quality of life 6 months after stroke: the role of anxiety and upper limb dysfunction. *Disabil Rehabil*. 2013;35(4):291–9.
62. Tyson SF, Hanley M, Chillala J, Selley AB, Tallis RC. Sensory loss in hospital-admitted people with stroke: characteristics, associated factors, and relationship with function. *Neurorehabil Neural Repair*. 2008;22(2):166–72.
63. Wade DT, Langton-Hewer R, Wood VA, Skilbeck CE, Ismail HM. The hemiplegic arm after stroke: Measurement and recovery. *J Neurol Neurosurg Psychiatry*. 1983;46:521–4.
64. Chester CS, McLaren CE. Somatosensory evoked response and recovery from stroke. *Arch Phys Med Rehabil*. 1989;70(7):520–5.
65. Vanbellingen T, Kersten B, Van de Winckel A, Bellion M, Baronti F, Müri R, Bohlhalter S. A new bedside test of gestures in stroke: the apraxia screen of TULIA (AST). *J Neurol Neurosurg Psychiatry*. 2011;82(4):389–92.
66. Scott SH. Apparatus for measuring and perturbing shoulder and elbow joint positions and torques during reaching. *J Neurosci Methods*. 1999;89(2):119–27.
67. Semrau JA, Herter TM, Scott SH, Dukelow SP. Robotic identification of kinesthetic deficits after stroke. *Stroke*. 2013;44(12):3414–21.
68. Mochizuki G, Centen A, Resnick M, Lowrey C, Dukelow SP, Scott SH. Movement kinematics and proprioception in post-stroke spasticity: assessment using the Kinarm robotic exoskeleton. *J Neuroeng Rehabil*. 2019;16(1):1–13.
69. Wood MD, Simmatis LE, Jacobson JA, Dukelow SP, Boyd JG, Scott SH. Principal components analysis using data collected from healthy individuals on two robotic assessment platforms yields similar behavioral patterns. *Front Human Neuroscience*. 2021;15:218.
70. Dexterit-E Software: <https://kinarm.com/kinarm-products/dexterit-e/>. Last visited on February 07, 2022.
71. Kinarm Lab: <https://kinarm.com/kinarm-products/kinarm-exoskeleton-lab/>. Last visited on December 16, 2021.
72. Box GE, Cox DR. An analysis of transformations. *J Roy Stat Soc: Ser B (Methodol)*. 1964;26(2):211–43.
73. Pearson ES, Please NW. Relation between the shape of population distribution and the robustness of four simple test statistics. *Biometrika*. 1975;62(2):223–41.
74. Head H, Holmes G. Sensory disturbances from cerebral lesions. *Brain*. 1911;34(2–3):102–254.
75. Fukutake T, Hirayama K, Komatsu T. Transient unilateral catalepsy and right parietal damage. *Psychiatry Clin Neurosci*. 1993;47(3):647–50.
76. Hiraga A, Sakakibara R, Mizobuchi K, Asahina M, Kuwabara S, Hayashi Y, Hattori T. Putaminal hemorrhage disrupts thalamocortical projection to secondary somatosensory cortex: case report. *J Neurosci*. 2005;23(11–2):81–3.
77. Ithori N, Kawamura M, Araki S, Kawachi J. Kinesthetic alexia due to left parietal lobe lesions. *Eur Neurol*. 2002;48(2):87–96.
78. Kwakkel G, Kollen BJ, van der Grond J, Prevo AJ. Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. *Stroke*. 2003;34(9):2181–6.
79. Rand D, Weiss PL, Gottlieb D. Does proprioceptive loss influence recovery of the upper extremity after stroke? *Neurorehabil Neural Repair*. 1999;13(1):15–21.
80. Rand D, Gottlieb D, Weiss PL. Recovery of patients with a combined motor and proprioception deficit during the first six weeks of post stroke rehabilitation. *Phys Occup Ther Geriatr*. 2001;18(3):69–87.
81. Yoshida H, Kondo T, Nakasato N. Neuromagnetic investigation of somatosensory cortical reorganization in hemiplegic patients after thalamic hemorrhage. *J Phys Ther Sci*. 2008;20(2):123–7.
82. Sakakibara R, Kishi M, Ogawa E, Shirai K. Isolated facio-lingual hypoalgesia and weakness after a hemorrhagic infarct localized at the contralateral operculum. *J Neurol Sci*. 2009;276(1–2):193–5.
83. Tiffin J, Asher EJ. The Purdue Pegboard: norms and studies of reliability and validity. *J Appl Psychol*. 1948;32(3):234.
84. Gowland C, Stratford P, Ward M, Moreland J, Torresin W, Van Hullenaar S, Sanford J, Barreca S, Vanspall B, Plews N. Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. *Stroke*. 1993;24(1):58–63.
85. Twitchell TE. The restoration of motor function following hemiplegia in man. *Brain*. 1951;74(4):443–80.
86. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil*. 1987;1:6–18.
87. Van Rossum G, Drake FL. Python 3 Reference Manual, Scotts Valley, CA: CreateSpace. 2009.
88. Kleinbaum DG, Dietz K, Gail M, Klein M, Klein M. Logistic regression. New York: Springer-Verlag; 2002. p. 536.
89. Song YY, Ying LU. Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*. 2015;27(2):130.
90. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
91. Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2019;9(3): e1301.
92. Noble WS. What is a support vector machine? *Nat Biotechnol*. 2006;24(12):1565–7.
93. Bengio Y. Learning deep architectures for AI. Now Publishers Inc. 2009.
94. Mahesh B. Machine learning algorithms-a review. *Int J Sci Res (IJSR)*. 2020;9:381–6.
95. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Computer Science*. 2021;2(3):1–21.
96. Ayodele TO. Types of machine learning algorithms. *New Adv Machine Learning*. 2010;3:19–48.
97. Singh A, Thakur N, Sharma A. A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) 2016*; pp. 1310–1315. IEEE.
98. Kumar M, Thenmozhi M. Forecasting stock index movement: a comparison of support vector machines and random forest. In *Indian institute of capital markets 9th capital markets conference paper*. 2006.
99. Roßbach P. Neural networks vs random forests—does it always have to be deep learning. Germany: Frankfurt School of Finance and Management; 2018.
100. König G, Molnar C, Bischl B, Grosse-Wenttrup M. Relative feature importance. In *2020 25th International Conference on Pattern Recognition (ICPR) 2021*; pp. 9318–9325. IEEE.
101. Choudhary P, Kramer A, c. datascience.com team. Skater: Model interpretation library. 2018. Available: <https://oracle.github.io/Skater/reference/interpretation.html?> Last Visited on 4 March 2022.
102. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: machine learning in Python. *J Machine Learning Res*. 2011;12:2825–30.
103. Veale JF. Edinburgh handedness inventory—short form: a revised version based on confirmatory factor analysis. Laterality Asymmetries Body Brain Cognition. 2014;19(2):164–77.

104. Confusion Matrix, Accuracy, Precision, Recall, F1 Score: <https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd>. Last Visited on February 08, 2022.
105. Goble DJ, Lewis CA, Brown SH. Upper limb asymmetries in the utilization of proprioceptive feedback. *Exp Brain Res*. 2006;168(1–2):307–11.
106. Findlater SE, Desai JA, Semrau JA, Kenzie JM, Rorden C, Herter TM, Scott SH, Dukelow SP. Central perception of position sense involves a distributed neural network—evidence from lesion-behavior analyses. *Cortex*. 2016;79:42–56.
107. Chilvers MJ, Hawe RL, Scott SH, Dukelow SP. Investigating the neuroanatomy underlying proprioception using a stroke model. *J Neurol Sci*. 2021;430: 120029.
108. Dextertit-E Explorer User Guide 3.9: <https://kinarm.com/download/dextertit-e-explorer-3-9-user-guide/>. Last visited on December 16, 2021.
109. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. *PLoS ONE*. 2017;12(5): e0177726.
110. Gromiha MM, Suresh MX. Discrimination of mesophilic and thermophilic proteins using machine learning algorithms. *Proteins Struct Funct Bioinform*. 2008;70(4):1274–9.
111. Govindarajan P, Soundarapandian RK, Gandomi AH, Patan R, Jayaraman P, Manikandan R. Classification of stroke disease using machine learning algorithms. *Neural Comput Appl*. 2020;32(3):817–28.
112. Hung CY, Chen WC, Lai PT, Lin CH, Lee CC. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2017*; (pp. 3110–3113). IEEE.
113. Chang SC, Chu CL, Chen CK, Chang HN, Wong AM, Chen YP, Pei YC. The comparison and interpretation of machine-learning models in post-stroke functional outcome prediction. *Diagnostics*. 2021;11(10):1784.
114. Semrau JA, Wang JC, Herter TM, Scott SH, Dukelow SP. Relationship between visuospatial neglect and kinesthetic deficits after stroke. *Neurorehabil Neural Repair*. 2015;29(4):318–28.
115. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol*. 2014;14(1):1–13.
116. Lin WY, Chen CH, Tseng YJ, Tsai YT, Chang CY, Wang HY, Chen CK. Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation. *Int J Med Informatics*. 2018;111:159–64.
117. Ge Y, Wang Q, Wang L, Wu H, Peng C, Wang J, Xu Y, Xiong G, Zhang Y, Yi Y. Predicting post-stroke pneumonia using deep neural network approaches. *Int J Med Informatics*. 2019;132: 103986.
118. Li X, Wu M, Sun C, Zhao Z, Wang F, Zheng X, Ge W, Zhou J, Zou J. Using machine learning to predict stroke-associated pneumonia in Chinese acute ischaemic stroke patients. *Eur J Neurol*. 2020;27(8):1656–63.
119. Rondina JM, Filippone M, Girolami M, Ward NS. Decoding post-stroke motor function from structural brain imaging. *NeuroImage Clinical*. 2016;12:372–80.
120. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, Wang Y, Douiri A, Wolfe CD, Bray B. A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLoS ONE*. 2020;15(6): e0234722.
121. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics*. 2017;37(2):505–15.
122. Magoulas GD, Prentza A. Machine learning in medical applications. In *Advanced course on artificial intelligence* (pp. 300–307). Springer, Berlin, Heidelberg. 1999.
123. De Bruijne M. Machine learning approaches in medical image analysis: from detection to diagnosis. 2016.
124. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. 2019;29(2):102–27.
125. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, Zhao S. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discovery*. 2019;18(6):463–77.
126. Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discovery Today*. 2015;20(3):318–31.
127. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discovery Today*. 2021;26(1):80.
128. Myszczyńska MA, Ojames PN, Lacoste AM, Neil D, Saffari A, Mead R, Hautbergue GM, Holbrook JD, Ferraiuolo L. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat Rev Neurol*. 2020;16(8):440–56.
129. Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. *J Personalized Med*. 2020;10(2):21.
130. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. Covid-19 outbreak prediction with machine learning. *Algorithms*. 2020;13(10):249.
131. Kim J, Ahn I. Infectious disease outbreak prediction using media articles with machine learning models. *Sci Rep*. 2021;11(1):1–13.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

