Indian J Physiol Pharmacol 2017; 61(4) : 323–332

Original Article

Autonomic Function Based Classification of Spinocerebellar Ataxia Type 1 and 2 Using Machine Learning Classifiers

Dibashree Tamuli[1], Anoop Kant Godiyal[3], Manpreet Kaur[1], Ashok Kumar Jaryal[1], Achal Kumar Srivastava[2] and Kishore Kumar Deepak[1]*

Departments of Physiology[1] and Neurology[2],
All India Institute of Medical Sciences,
New Delhi, India and

Centre for Biomedical Engineering[3],
Indian Institute of Technology,
New Delhi, India

Abstract

Spinocerebellar ataxia (SCA) is a progressive neurodegenerative disorder characterized by autonomic dysfunction. SCA has multiple genetically classified subtypes, amongst which SCA1 and SCA2 are most prevalent in India. Autonomic function based characterization of SCA patients into respective subtypes has not been done. We have evaluated autonomic function - heart rate variability (HRV), systolic blood pressure variability (BPV), systolic baroreflex sensitivity (BRS) and composite autonomic severity score (CASS) in SCA patients (SCA1 = 31; SCA2 = 40). To evaluate the classification performance of the battery of autonomic function tests (AFT), linear discriminant analysis (LDA) and support vector machine (SVM) classifiers were used. The average classification accuracy for SCA subtypes were 80% by LDA and 70% by SVM. Interestingly, individually the autonomic function tests do not differ between SCA1 and SCA2 but when they are used together by classifier - a conclusive pattern to characterize the SCA subtypes emerges. This is the first study to classify SCA patients into their respective subtypes using a novel machine learning approach on their autonomic function profile.

Introduction

Spinocerebellar ataxia (SCA) is a heterogeneous group of progressive neurodegenerative disorder of cerebellum, its afferent and/or efferent connections. More than 30 distinct genetic subtypes of SCA are known (1). Amongst them, SCA1 and SCA2 are most prevalent in India (2, 3). Over the years, clinicians have been faced with immense difficulty in classifying patients into SCA1 and SCA2 based on clinical signs and symptoms. The only existing gold standard for classification is based on genetic analysis quantifying the number of CAG trinucleotide repeats (4, 5). The

***Corresponding author :**

Prof. KK Deepak, Room no. 2009, Department of Physiology, Teaching block, Second Floor, New Delhi – 110 029.
Email: kkdeepak@gmail.com; Mob: 8826168442/9868397129

(Received on August 1, 2017)

problem with genetic analysis lies in it being expensive, limited centers for testing and cumbersome procedure requiring trained manpower. There is an existing need for non-invasive, cost-effective and reliable biomarkers to classify SCA1 and SCA2 patients.

Autonomic dysfunction is a predominant manifestation in SCA patients (6-10). Thus a classification based on autonomic function could serve as a promising biomarker. Therefore, we have studied autonomic function tests - heart rate variability (HRV), systolic blood pressure variability (BPV), systolic baroreflex sensitivity (BRS) and composite autonomic severity score (CASS) in SCA patients. HRV quantifies the beat to beat variation in RR interval and is a validated measure of cardiac autonomic tone (11-14). Similarly, BPV measures the variation in beat to beat blood pressure and signifies the autonomic regulation of blood pressure (13, 14). BRS is a composite measure of the overall cardiovascular autonomic regulation (15). CASS was developed as a single validated score for laboratory quantification of autonomic failure (16, 17). This score standardizes for known confounding factors like age and gender.

The idea was to derive a characterization of SCA1 and SCA2 patients on the basis of their autonomic function profile. The autonomic function tests (HRV, BPV, BRS and CASS) were assessed in SCA patients (SCA1 = 31; SCA2 = 40) and we tried to classify them into SCA subtypes using machine learning approach. Machine learning is an emerging field where algorithms are made which can learn on the basis of data and subsequently can be used to classify data (18). These algorithms can be used for classification and enable reliable and repeatable decisions. We have used LDA and SVM - two of the machine learning algorithms. LDA is a one of the pattern recognition algorithms of machine learning used to find a linear combination of features that characterizes or separates two or more subsets (19-21). SVM is another algorithm of machine learning which can be used to create decision boundary to decipher two or more groups using both linear and non-linear approaches (22).

This paper presents a study to explore a novel machine learning approach using classifiers based on autonomic function to classify SCA1 and SCA2.

Methods

A. Patients

The study was conducted on genetically confirmed SCA1 (n=31, age = 35.8±7.8 years) and SCA2 (n=40, age = 33.3±10.5 years) recruited from outpatient department of Neurology at the All India Institute of Medical Sciences (AIIMS), New Delhi. The patients had no other co-morbidities (vascular disease, heart disease, respiratory problems etc.) and were not on any medications known to affect their autonomic function. Informed written consent was obtained from all patients. The study protocol was approved by Institute Ethics Committee, All India Institute of Medical Sciences (AIIMS), New Delhi.

B. Study protocol

All subjects were instructed to refrain from food and caffeine intake 4-h before the recording. Patients were given 15 minutes of supine rest before assessment of autonomic function parameters in ambient temperature (25-27°C) maintained room. The measurements of heart rate variability (HRV), blood pressure variability (BPV), baroreflex sensitivity (BRS), autonomic reactivity tests and quantitative sudomotor axon reflex test (QSART) were performed at the same visit in morning hours. Arterial pressure signal at sampling frequency of 200 Hz was measured beat-to-beat non-invasively with finger photoplethysmography using Finometer® model 2 (FMS, Finapres Medical Systems, Amsterdam, the Netherlands) based on the volume clamp method of Penaz along with Lead II ECG at 1 kHz for 5 min under spontaneous breathing by patient using digital band pass filter with low cut-off frequency of 0.5 Hz and high cut-off frequency of 35 Hz (Fig. 1). Autonomic function tests comprised of autonomic tone (HRV, BPV, BRS) and autonomic reactivity tests which were quantified using composite autonomic severity score (CASS).

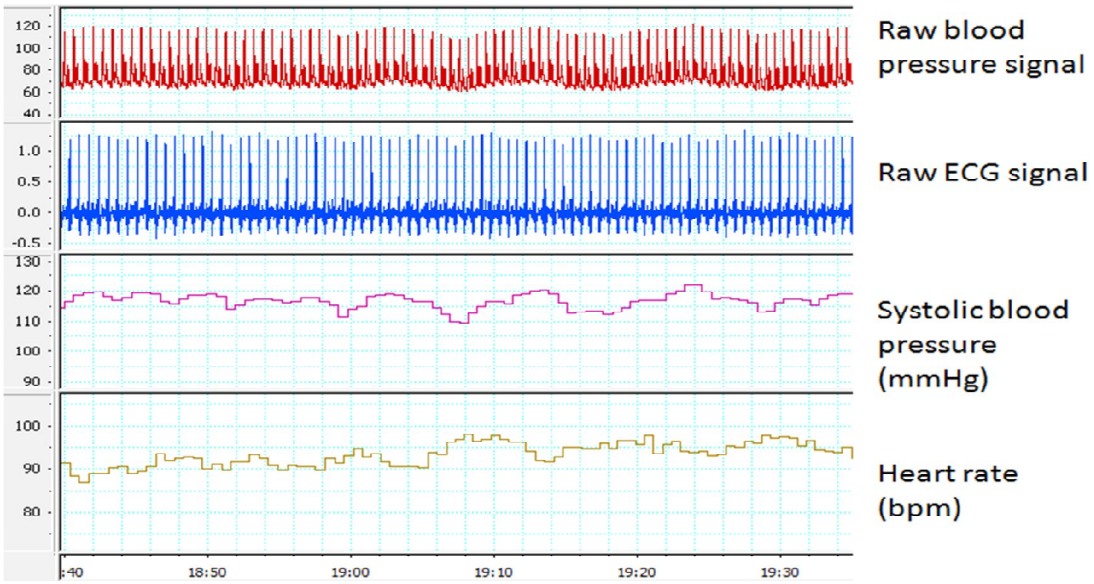

Fig. 1 : This figure represents the record of blood pressure and lead II ECG in a patient of spinocerebellar ataxia during spontaneous breathing. Systolic blood pressure and heart rate signals are derived from raw blood pressure and ECG signals respectively.

C. Heart rate and blood pressure signal analysis

Analysis of all recorded signals was done offline using the Nevrokard HRV and systolic BPV analysis software for calculation of the time and frequency domain parameters along with the derivation of spontaneous systolic BRS.

1) Heart rate variability (HRV)

ECG was analysed offline using Labchart Pro 7 (AD instruments, Australia) and HRV was computed using Nevrokard software (Nevrokard Kiauta, Slovenia) on beat-to-beat series. HRV was assessed by time domain, frequency domain and non-linear analysis.

a) *Time domain analysis of heart rate variability:* The time domain indices of HRV computed were SDNN (Standard deviation of all NN intervals) and RMSSD (The square root of the mean of the sum of the squares of differences between adjacent NN intervals) (11).

b) *Frequency domain analysis of heart rate variability:* The interval between successive R waves was computed (RRI) from the original electrocardiographic signal resampled at 7 Hz. The R-R tachogram was further analysed by power spectral density (PSD) analysis of the R-R sequences. The PSD was calculated using fast Fourier transformation (FFT) using Hanning window in the three frequency bands such as very low frequency (VLF; frequency 0.001–0.04 Hz), LF (frequency 0.04–0.15 Hz) and HF (frequency 0.15–0.40 Hz) (11). The normalised LF, HF and total power components were calculated.

c) *Non-linear analysis of heart rate variability:* HRV was also calculated by non-linear approach using a Poincare plot. To quantify the plot's geometric appearance, the technique of fitting an ellipse to the plot was used from which two indices were obtained - SD1(the standard deviation of instantaneous beat-to-beat R-R interval variability or minor axis of the ellipse) and SD2 (the standard deviation of long term R-R interval variability or major axis of the ellipse) (12). The normalised SD1 and SD2 were determined.

2) Systolic blood pressure variability (BPV)

Systolic BPV was also determined by time and frequency domain analysis. The systolic BPV was analysed in time domain to calculate SDNN and RMSSD. For spectral analysis, PSD was calculated using FFT using Hanning window in the three

frequency bands such as VLF (frequency 0.001–0.04 Hz), LF (frequency 0.04–0.15 Hz) and HF (frequency 0.15–0.40 Hz). PSDs were plotted in $mmHg^2$/Hz for systolic BPV (13,14).

3) Systolic baroreflex sensitivity (BRS)

The quantification of systolic BRS was done by sequence and spectral methods as described earlier (14,15). To detect the sequences, the criteria were (i) R-R variation >5 ms, (ii) BP changes >0.5 mmHg, (iii) sequences >3 beats and (iv) sequences correlation coefficient >0.85. The slope of regression line between R-R interval and systolic blood pressure (SBP) gives estimates of the BRS. In the spectral method, analysis of R-R interval and simultaneously recorded beat-to-beat blood pressure was done. Thus, baroreflex gain was computed by dividing the amplitude of R-R oscillations in low frequency (LF) band (0.04–0.15 Hz) and high frequency (HF) band (0.15–0.4 Hz) by the amplitude of corresponding oscillations in SBP. A Hanning window was used and magnitude of squared coherence (K^2) function between the SBP and R-R interval was computed for the calculation of gain in the transfer function in each frequency band. The coherence between the R-R interval and beat-to-beat SBP was assessed by cross-spectral analysis. The α-index (gain in relationship between the R-R intervals and beat-to-beat SBP) was calculated only when the K^2 between the RRI spectrum and the SBP spectrum >0.5. α-LF is K^2 function between pressure and R-R interval in LF band, whereas α-HF is K^2 function between the SBP and R-R interval in an HF band. Both α-LF and α-HF are expressed as ms/mmHg.

4) Composite autonomic severity score (CASS) analysis

To quantify autonomic failure a single validated score, CASS was developed which standardizes for the confounding effects of age and gender. CASS ranges from 0 to 10 and constitutes of cardiovagal (0–3), adrenergic (0–4) and sudomotor (0–3) subscores (Table I) (16,17). Cardiovagal subscore comprises of results of deep breathing test (DBT) and Valsalva maneuver (VM), quantifying the heart rate changes during these respective maneuvers. On the other hand, adrenergic subscore constitutes of blood pressure changes during VM and head up tilt (HUT). Sudomotor autonomic function was assessed by QSART. In QSART, sudomotor responses were recorded from four sites of right side: medial forearm, proximal lateral leg, medial distal leg and proximal foot by placing four capsules (Fig. 2). The stimulus was iontophoresis of 10% acetylcholine solution for 5 min, and the responses were recorded in a sweat cell using transdermal self adhesive electrodes (Fig. 3).

TABLE I: The composite autonomic severity score (CASS).

Cardiovagal subscore	
1	HR_{DB} or VR reduced but ≥50% minimal normal value
2	HR_{DB} or VR reduced to <50% of minimal normal value
3	Both HR_{DB} and VR reduced to <50% of minimal normal value
Adrenergic subscore	
1	Phase II_e reduction <40 but >20 mmHg mean BP or Phase II_l does not return to baseline or Pulse pressure reduction ≥50% of baseline or Systolic BP recovery time 4-5 seconds
2	Phase II_e reduction >40 >20 mmHg mean BP + Phase II_l or IV absent or Systolic BP recovery time 6-9 seconds
3	Orthostatic hypotension (systolic BP reduction = 30 mmHg; mean BP reduction = 20 mmHg) or Phase II_e decrease of >40 mmHg + Absent phases II_l + IV or Absent phases II_l + IV + Systolic BP recovery time = 10 seconds
4	Criteria for 3 + Orthostatic hypotension (systolic BP reduction = 30 mmHg; mean BP reduction = 20 mmHg)
Sudomotor subscore	
1	Single site abnormal and ≥50% of the lower limit
2	Single site <50% of the lower limit or two or more sites reduced and ≥50% of the lower limit
3	Two or more sites <50% of lower limit

*This criteria is based on Low PA (Ref. 17) and Novak P (Ref. 16). [0 in each of these subscores denotes normal response].
Abbreviations: BP = blood pressure; HR_{DB} = heart rate response to deep breathing; VR = Valsalva ratio; II_e = phase 2 early; II_L = phase 2 late.

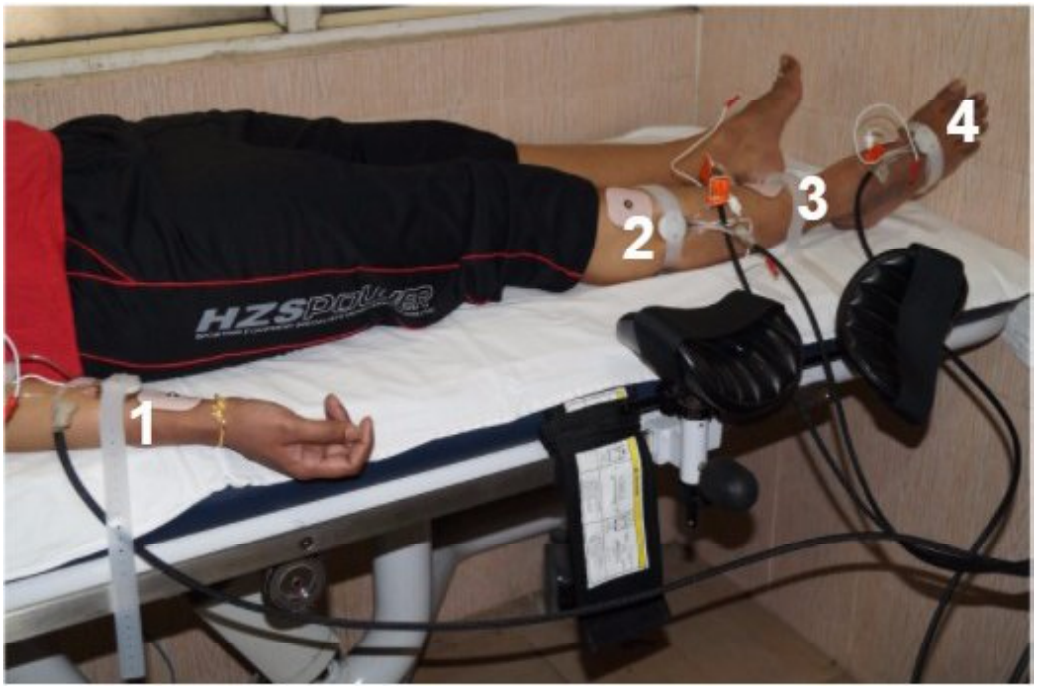

Fig. 2 : This figure depicts the placement of electrodes in quantitative sweat axon reflex test from four sites of right side: medial forearm, proximal lateral leg, medial distal leg and proximal foot by placing four capsules.

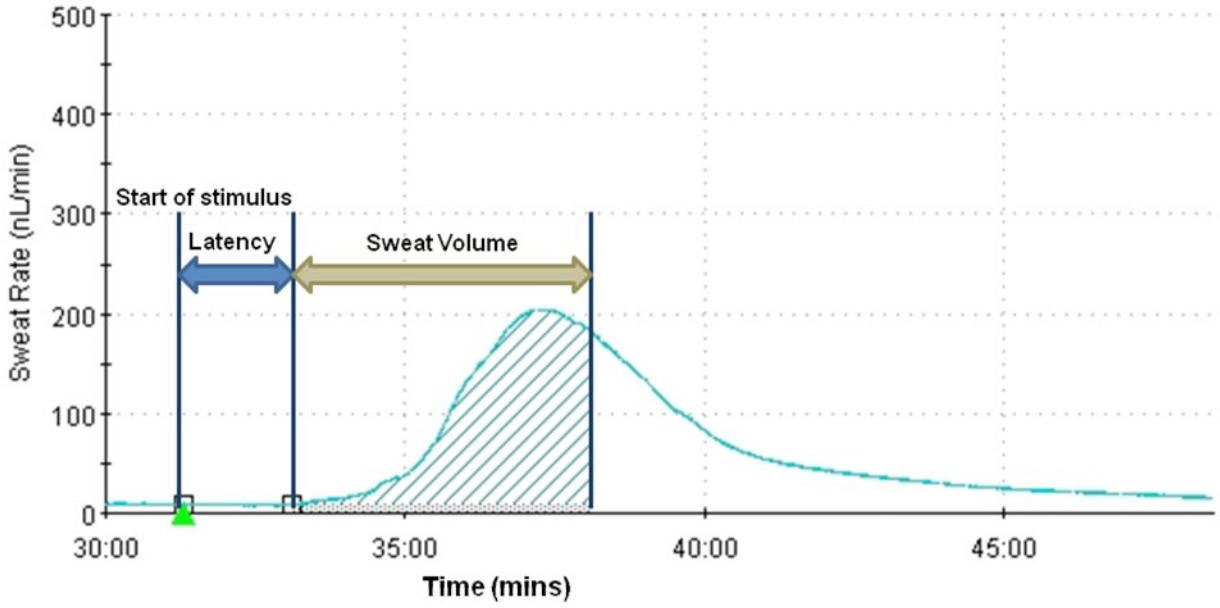

Fig. 3 : This figure shows the record obtained in quantitative sweat axon reflex test at right medial forearm site representing the two parameters - latency and sweat volume (calculated by area under the curve and used for composite autonomic severity score calculation).

D. Statistical analysis

Each parameter was tested for distribution of the data based on standard normality tests (D' Agostino-Pearson omnibus normality test and Shapiro–Wilk test). In case of Gaussian data distribution, parametric tests were applied and for non-Gaussian distribution, appropriate non-parametric tests were applied. The following tests were used: t-test for normally distributed data and the Mann-Whitney U

test for non-Gaussian distribution. The categorical variables were compared by using Fisher's exact test. The statistical analyses were performed using Graph-Pad Prism Version 5.00 for Windows (GraphPad Software, Inc., USA) and analysis of CASS was done by using MATLAB (R2015a).

E. Classifier Design

1) Linear Discriminant Analysis (LDA)

LDA classifier has an inherent advantage of being computationally efficient, easy to implement and having good real time performance (19). Especially, small data set can also be utilized by this classifier to get a reliable output (21). It provides an adequate separation between the given classes by looking into the linear combinations of input variables. In a LDA, if S_i be the data element having m features and n be the number of subjects in an experiment, each subject can be assigned to any of the two categories € = {0, 1}. Then S renders in a matrix of size n×m and € in a vector of size n×1. Classes 0 and 1 have the number of elements N_0 and N_1 respectively.

The mean of each class € over all S_i is denoted by μ_c where i elements in class €. The overall mean μ of the data is

$$\mu = \frac{N_0\mu_0 + N_1\mu_1}{N_0 + N_1}$$

For this the expectation value of data is expressed as covariance matrix C.

$$C = E\left((S-\mu)^T(S-\mu)\right)$$

The weight vector w and the offset w_0 can be expressed as:

$$w = C^{-1}(\mu_1 - \mu_0)^T \text{ and } w_0 = -\mu w$$

A separating hyperplane is determined by the weight vector w in m dimensional feature space. Any element x can be found at a normal distance D(x):

$$D(x) = xw + w_0$$

Any element x is assigned to class 1 if D(x) is greater than zero and x is assigned to class 0 if D(x) is less than zero. All elements are part of the separating hyperplane if D(x) equals to zero.

2) Support Vector Machine (SVM)

SVM classifiers are robust for condensing information in the learning data and offer a good performance of classification even with smaller data points. SVM is implemented by constructing a hyperplane as a decision surface so as the margin of separation between the positive and a negative example is maximized. It works with the higher dimensional feature space formed by the nonlinear mapping, $\varphi(x)$ of the n-dimensional input vector into a K-dimensional feature space (22). The equation of the hyperplane separating two different classes is given by the relation:

$$y(x) = W^T\varphi(X) = \sum_{j=1}^{K}\omega_j\varphi_j(x) + \omega_0$$

with $W = [\omega_0, \omega_1, !\ , \omega_k]^T$ is the weight vector of the network. A non-linear (polynomial) kernel was used for the SVM classifier; $k(x_i, x_j) = (x_i^T x_j + c)^d$, where x_i and x_j are the feature vectors of i^{th} and j^{th} classes, respectively, $c \geq 0$ is a trading off parameter for influencing higher-order versus lower-order terms in the polynomial, T denotes transpose and d denotes the degree of polynomial.

Previously both the classifiers have been extensively used for the pattern recognition and are computationally efficient. A total of 19 AFT features were feed into the classifiers to determine the overall performance. The data was randomly sampled at a train to test ratio of 85:15 for both of the classifiers. The training data was used to train the system and testing data was analyzed to test the performance

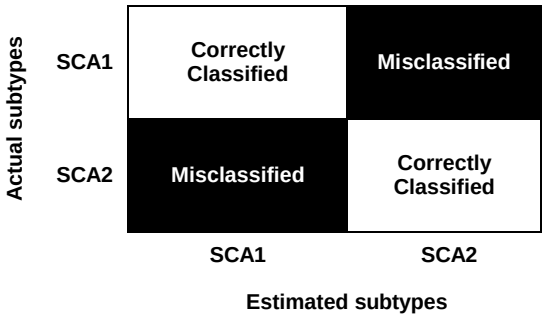

Fig. 4 : Concept of 2x2 confusion matrix with the actual condition representing the genetically known spinocerebellar ataxia subtypes and estimated condition representing the ones classified by either linear discriminant analysis or support vector machine.

of the system. To check the performance of the system, a confusion matrix (Fig. 4) was made - where the diagonal element represents the correct classification and the non-diagonal element shows the misclassification. Three parameters - sensitivity, specificity and overall classification accuracy were calculated for the data sets. Sensitivity and specificity for both SCA1 and SCA2 were calculated using previously described formulas (23).

The overall classification accuracy (CA) was calculated as :

$$CA = \frac{\text{Number of correctly classified trials}}{\text{Total number of trials}} \times 100$$

Results

a) Demographic and cardio-respiratory profile of SCA1 and SCA2 patients

The two SCA subtypes were comparable in demographic and basal cardio-respiratory parameters as shown in the Table II.

b) Autonomic function analysis by heart rate variability (HRV), systolic blood pressure variability (BPV), systolic baroreflex sensitivity (BRS) and composite autonomic severity score (CASS)

The time and frequency domain measures of HRV (linear and non-linear), systolic BPV and systolic BRS and CASS in the patients of SCA1 and SCA2 are presented in Table III. There was no significant difference in all these parameters between SCA1 and SCA2 patients.

TABLE II : Demographic and cardio-respiratory profile of the patients.

Parameter	*SCA1 (n=31)*	*SCA2 (n=40)*	*p-value*
Male: Female (n)	24:7	30:10	0.999
Age (years)	35.3±7.8	33.3±10.5	0.368
Age of onset (years)	30.5±7.5	27.1±9.6	0.111
Disease duration (years)#	5.0 (2.0–6.0)	6.0 (4.0–7.8)	0.056
BMI (kg/m^2)	24.2±3.0	23.7±3.3	0.461
HR (bpm)	81.5±13.3	81.4±11.7	0.990
SBP (mmHg)	120.6±14.9	120.2±12.2	0.897
DBP (mmHg)#	71.8 (61.1–85.0)	70.6 (57.8–80.1)	0.388
MBP (mmHg)#	89.1 (76.3–96.3)	87.7 (75.2–93.9)	0.351
Respiratory rate (breaths/min)	20.3±4.1	19.2±3.9	0.221

All values are expressed as Mean±SD except for the parameters denoted by # [values are expressed as median (interquartile range)].

Abbreviation: SCA = spinocerebellar ataxia; n = number, SBP = systolic blood pressure; DBP = diastolic blood pressure; MBP = mean blood pressure; BMI = body mass index; HR = heart rate; bpm = beats per minute.

c) Classification of SCA1 and SCA2 by autonomic function tests using linear discriminant analysis (LDA) and support vector machine (SVM) classifier

The parameters measured by the autonomic function tests (19 in number) as shown in Table III were feed into the two classifiers i.e. LDA and SVM and the overall performance was determined. A confusion matrix was made representing correct classification and misclassification for both the groups SCA1 and SCA2. The confusion matrix for LDA classifier has been depicted in Fig. 5. LDA exhibited 80% overall classification accuracy with sensitivity and specificity of 80% for both SCA1 and SCA2 subtypes respectively. On the other hand, a lower accuracy was seen in the confusion matrix for

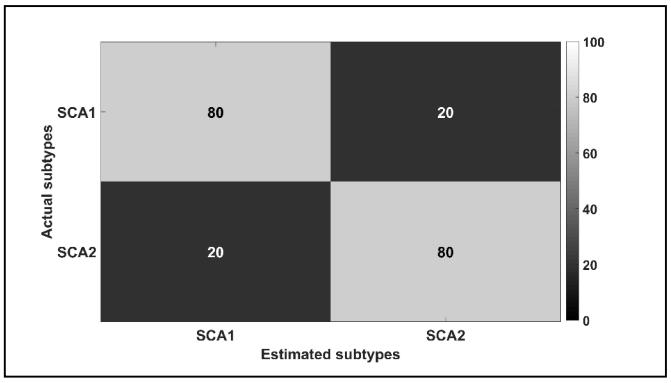

Fig. 5 : Confusion matrix for spinocerebellar ataxia type 1 (SCA1) and spinocerebellar ataxia type 2 (SCA2) using linear discriminant analysis.

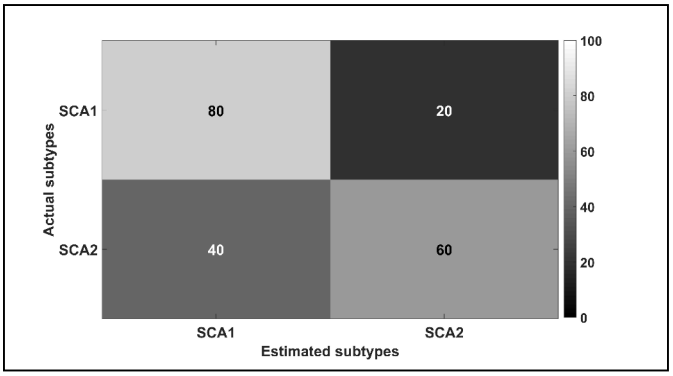

Fig. 6 : Confusion matrix for spinocerebellar ataxia type 1 (SCA1) and spinocerebellar ataxia type 2 (SCA2) using support vector machine.

TABLE III : Time and frequency domain measures of HRV (linear and non-linear), systolic BPV, systolic BRS and CASS in SCA1 and SCA2 patients.

Parameter	*SCA1 (n=31)*	*SCA2 (n=40)*	*p-value*
Time domain HRV indices (linear)			
SDNN (ms) #	23.59±10.75	24.07±11.71	0.859
RMSSD (ms)	13.19 (9.30–17.18)	13.57 (8.17–21.98)	0.912
Frequency domain HRV indices (linear)			
LF (nu)	62.44 (47.65–69.64)	58.74 (45.44–70.49)	0.931
HF (nu)	30.30 (23.10–41.19)	28.57 (20.22–44.68)	0.551
Total Power (nu)	196.60 (148.40–273.40)	170.80 (146.90–216.20)	0.177
HRV indices (non-linear)			
SD1 (nu)	1.24 (0.92–1.75)	1.31 (0.79–1.83)	0.995
SD2 (nu) #	4.07±1.56	4.2±1.74	0.912
Time domain systolic BPV indices			
SDNN (mmHg) #	5.17±2.13	5.34±2.26	0.743
RMSSD (mmHg)	1.92 (1.63–2.32)	2.00 (1.56–2.44)	0.689
Frequency domain systolic BPV indices			
LF (nu)	74.29 (61.74–83.73)	73.06 (58.78–83.92)	0.894
HF (nu)	21.08 (10.04–33.53)	18.25 (9.04–30.05)	0.498
Power (nu)	241.30 (192.20–501.60)	239.50 (164.70–318.10)	0.196
Time domain systolic BRS indices			
All BRS (ms/mmHg)	7.43 (5.54–9.87)	7.92 (5.48–12.24)	0.693
Frequency domain systolic BRS indices			
α-LF (ms^2)	1.40 (1.17–2.57)	1.47 (0.94–2.52)	0.469
α-HF (ms^2)	1.98 (1.44–4.18)	2.05 (1.05–3.88)	0.741
Composite autonomic severity score (CASS)			
a) Cardiovagal subscore	0 (0–1)	0.5 (0–2)	0.397
b) Adrenergic subscore	1 (1–1)	1 (1–1)	0.364
c) Sudomotor subscore	3 (2–3)	3 (2–3)	0.935
Total CASS	4 (4–6)	5 (4–5)	0.904

All values are expressed as median (interquartile range).
#The values are expressed as Mean±SD, n denotes number.
Abbreviations: SCA = spinocerebellar ataxia; HRV = heart rate variability; SDNN = standard deviation of all NN intervals; RMSSD = the square root of the mean of the sum of the squares of differences between adjacent NN intervals; LF = low frequency; HF = high frequency; SD = standard deviation; ms = millisecond; nu = normalized unit; BPV = blood pressure variability; BRS = baroreflex sensitivity.

the SVM classification for SCA patients (Fig. 6). For SVM classifier, SCA1 had 80% sensitivity and 60% specificity while SCA2 exhibited 60% sensitivity and 80% specificity. The overall accuracy of both the classifiers is shown in Fig. 7 which represents the average classification accuracy of both the methods - LDA and SVM in SCA1 and SCA2 patients together.

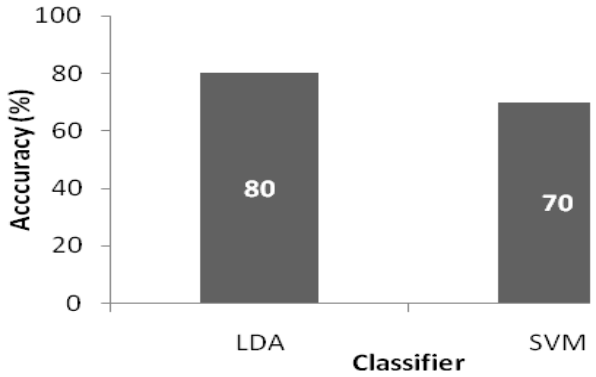

Fig. 7 : Average classification accuracy by using linear discriminant classifier (LDA) and support vector machine (SVM) for both spinocerebellar ataxia type 1 (SCA1) and spinocerebellar ataxia type 2 (SCA2) patients together.

Discussion

Almost 40-70% of Spinocerebellar Ataxia (SCA) patients have autonomic dysfunction (6-10). Autonomic function testing is feasible, non-invasive

and inexpensive technique for clinical assessment. SCA patients' classification into their respective subtypes is still an uphill task for clinicians. We hypothesized that certain autonomic profile pattern might typify certain ataxia subtypes. Thus this study was designed to attempt to characterize the autonomic functions in two most prevalent forms of SCA - SCA1 and SCA2 and then utilize them to classify SCA patients. Interestingly, individually these tests do not differ between SCA1 and SCA2 but when they are used together by classifier there emerges a definitive pattern to characterize the SCA subtypes. On similar lines, other authors have tried to classify SCA subtypes based on either their clinical symptoms or their magnetic resonance imaging (MRI) findings (24-26). But their limitations lies in the fact that they require expertise, involve expensive equipments and might be tedious for the patients.

This study used LDA and SVM classifier for classification of SCA patients. These classifier have been previously used in ataxia patients to characterize on the basis of either gait patterns or MRI findings (27, 28). The autonomic function based classifier in the present work could achieve an average accuracy of 80% using LDA classifier, whereas this accuracy was decreased to 70% by using SVM classifier. LDA emerges as a better classifier for our data with higher accuracy. The advantage of LDA is its capacity to precisely classify even with a smaller learning data and thus can be implemented faster with less complexity, which accounts for its better performance (21).

Our findings indicate that certain genetic subtypes of SCA tend towards a unique autonomic profile. Our study suggests that using autonomic function based parameters we can effectively classify SCA1 and SCA2 patients using LDA classifier.

These algorithms are freely available for use in MATLAB, R and many other platforms where the user can feed his data and utilize them to categorize it. The advantage of machine learning is it can be used to discover relevant features in otherwise disordered datasets and it can sort a large array of parameters together (which is not humanely possible) to categorize datasets accurately. It can effectively handle multi-dimensional huge set of data in less time and thus saves lots of time and resources. Machine learning tools have a huge scope in fields like physiology where we handle enormous amount of biological signal data having huge variability.

Our study was focused on SCA1 and SCA2 patients; however inclusion of other subtypes of SCA would have augmented the clinical utility. Because autonomic profiles are highly influenced by the patient population and case ascertainment, these results will be most valuable for the present patient population seen at this referral center. Despite these limitations, the present autonomic function tests offer promising, cost- effective and reliable classification of SCA1 and SCA2. The future course would include recruitment of more number of SCA patients.

This study evaluated full battery of autonomic function tests in order to classify SCA1 and SCA2 subtypes. Although alone these tests are unable to differentiate SCA1 and SCA2 but when utilized together using classifier they successfully categorize SCA subtypes. This system offers a good accuracy along with being non-invasive, cost effective and convenient. Thus, it may prove to be a clinically reliable, inexpensive and non-invasive alternative for the characterization of SCA1 and SCA2 patients.

Acknowledgements

We would like to acknowledge Mohammed Faruq, Institute of Genomics and Integrative Biology (IGIB), New Delhi for providing us the genetic testing reports for SCA1 and SCA2 patients in his laboratory, which was used to classify the subtypes.

References

1. Rüb U, Schöls L, Paulson H et al. Clinical features, neurogenetics and neuropathology of the polyglutamine spinocerebellar ataxias type 1, 2, 3, 6 and 7. *Prog Neurobiol* 2013; 104: 38–66.
2. Krishna N, Mohan S, Yashavantha BS et al. SCA 1, SCA 2 & SCA 3/MJD mutations in ataxia syndromes in southern

India. *Indian J Med Res* 2007; 126(5): 465–470.

3. Saleem Q, Choudhry S, Mukerji M et al. Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation. *Human Genetics* 2000; 106(2): 179–187.
4. Klockgether T, Paulson H. Milestones in ataxia. *Mov Disord Off J Mov Disord Soc* 2011; 26(6): 1134–1141.
5. Durr A. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *Lancet Neurol* 2010; 9(9): 885–894.
6. Singh KK, Jain S, Deepak KK, Kochhar KP, Maheshwari MC, Manchanda SK. Vasopressor responses in autosomal dominant cerebellar ataxia. *Indian J Physiol Pharmacol* 1996; 40(2): 145–150.
7. Netravathi M, Sathyaprabha TN, Jayalaxmi K, Datta P, Nirmala M, Pal PK. A comparative study of cardiac dysautonomia in autosomal dominant spinocerebellar ataxias and idiopathic sporadic ataxias. *Acta Neurol Scand* 2009; 120(3): 204–209.
8. De Joanna G, De Rosa A, Salvatore E et al. Autonomic nervous system abnormalities in spinocerebellar ataxia type 2: a cardiovascular neurophysiologic study. *J Neurol Sci* 2008; 275(1-2): 60–63.
9. Pradhan C, Yashavantha BS, Pal PK, Sathyaprabha TN. Spinocerebellar ataxias type 1, 2 and 3: a study of heart rate variability. *Acta Neurol Scand* 2008; 117(5): 337–342.
10. Montes-Brown J, Sánchez-Cruz G, García AM, Báez ME, Velázquez-Pérez L. Heart rate variability in type 2 spinocerebellar ataxia. *Acta Neurol Scand* 2010; 122(5): 329–335.
11. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation* 1996; 93(5): 1043–1065.
12. Tulppo MP, Makikallio TH, Takala TE, Seppanen TH, Huikuri HV. Quantitative beat-to-beat analysis of heart rate dynamics during exercise. American Journal of Physiology-Heart and Circulatory Physiology. 1996; 271(1): H244–H252.
13. Parati G, Saul JP, Di Rienzo M, Mancia G. Spectral analysis of blood pressure and heart rate variability in evaluating cardiovascular regulation. A critical appraisal. *Hypertens Dallas Tex* 1979. 1995; 25(6): 1276–1286.
14. Kaur M, Chandran D, Lal C et al. Renal transplantation normalizes baroreflex sensitivity through improvement in central arterial stiffness. *Nephrol Dial Transplant* 2013; 28(10): 2645–2655.
15. Parati G, Di Rienzo M, Mancia G. How to measure baroreflex sensitivity: from the cardiovascular laboratory to daily life. *J Hypertens* 2000; 18(1): 7–19.
16. Novak P. Quantitative autonomic testing. *J Vis Exp JoVE* 2011; (53): 2502.
17. Low PA. Composite autonomic scoring scale for laboratory quantification of generalized autonomic failure. *Mayo Clin Proc* 1993; 68(8): 748–752.
18. Witten IH, Frank E, Hall MA, Pal CJ., Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann Publishers, 2016.
19. Lachenbruch P.A., Discriminant Analysis. Hafner Press, New York 1975.
20. Yeh YC, Wang WJ, Chiou CW. Cardiac arrhythmia diagnosis method using linear discriminant analysis on ECG signals. *Measurement* 2009; 42(5): 778–789.
21. Sharma A, Paliwal KK. Linear discriminant analysis for the small sample size problem: an overview. *Interna Jour of Machi Learn and Cyberne* 2015; 6(3): 443–454.
22. Burges CJ. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov* 1998 Jun 1; 2(2): 121–167.
23. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain* 2008; 8(6): 221–213.
24. Maschke M, Oehlert G, Xie T-D, Perlman S, Subramony SH, Kumar N, et al. Clinical feature profile of spinocerebellar ataxia type 1-8 predicts genetically defined subtypes. *Mov Disord* 2005; 20(11): 1405–1412.
25. Ye C, Yang Z, Ying SH, Prince JL. Segmentation of the Cerebellar Peduncles Using a Random Forest Classifier and a Multi-object Geometric Deformable Model: Application to Spinocerebellar Ataxia Type 6. *Neuroinformatics* 2015; 13(3): 367–381.
26. Yang Z, Ye C, Bogovic JA et al. Automated cerebellar lobule segmentation with application to cerebellar structural analysis in cerebellar disease. *NeuroImage* 2016; 127: 435–444.
27. Pradhan C, Wuehr M, Akrami F et al. Automated classification of neurological disorders of gait using spatio-temporal gait parameters. *J Electromyogr Kinesiol* 2015; 25(2): 413–422.
28. Yang Z, Zhong S, Carass A, Ying SH, Prince JL. Deep Learning for Cerebellar Ataxia Classification and Functional Score Regression. Mach Learn Med Imaging, MLMI (Workshop). 2014; 8679: 68–76.