Predicting Humans' Balance Disorder Based on Center of Gravity Using Support Vector Machine

Tran Anh Vu[1], Hoang Quang Huy[1], Nguyen Viet Dung[1], Nguyen Phan Kien[1], Nguyen Thu Phuong[1,2], and Pham Thi Viet Huong[3(✉)]

[1] School of Electronics and Telecommunications, Hanoi University of Science and Technology, Hanoi, Vietnam
[2] National Geriatric Hospital, Hanoi, Vietnam
[3] International School, Vietnam National University, Hanoi, Vietnam
huongptv@isvnu.vn

Abstract. Currently, vestibular disorders are quite common in Vietnam. However, as far as we know, methods for vestibular diagnosis are only qualitative, which are mostly based on experiences and doctors' observations. Therefore, a demand for a quantitative method is needed to help doctors accurately diagnose the vestibular disease. Moreover, the method is expected to allow monitoring the patient's situation during the treatment. To response to this demand, this paper applied machine learning technique to build a model to predict a person who has balance disorder. The data is obtained by a self-made device to measure the Center of Gravity (CoG) from people with and without vestibular. Results show that our proposed quantitative method had high accuracy in predicting whether a certain person has balance disorder or not.

Keywords: Vestibular disorder · Center of Gravity (CoG) · Data analysis · SVM

1 Introduction

When a person has balance disorder, they may have the feeling of dizziness, unsteadiness, or a lightheaded status. They feel hard to keep balance or fall down when changing position. There are 4 main systems in the body that work together to ensure a good postural balance: inner ear (vestibular system), vision, muscle and joints, and sensory input [1]. When one of these systems does not work well, it interferes your life by making you uncomfortable and uncontrollably. Balance disorder can be the symptom of following diseases: benign paroxysmal positional vertigo (BPPV), Meniere's disease, Migraine, acoustic neuroma, vestibular neuritis, Ramsay Hunt syndrome, cardiovascular disease, vestibular problems [2]. To diagnose these diseases, doctors need to perform a complex process, including physical examination such as Romberg's sign, the rotary chair test, etc., and using medical equipment for Electronystagmography test and Videonystagmography test [3]. Hence, more advanced

P. Cong Vinh and A. Rakib (Eds.): ICCASA 2021, LNICST 409, pp. 38–47, 2021.
https://doi.org/10.1007/978-3-030-93179-7_3

medical techniques for better diagnosis are in high demand. A promising technique for determining balance disorder is to apply the Center of Gravity (CoG). The Center of Gravity (CoG) is an imaginary point around which the body's weight is evenly distributed. For patients with balance disorder, their CoG points oscillate around normal ones due to bad ability to keep balance, so their CoG patterns may be different.

In this paper, we use the CoG data, which is the measurement of people's CoG location, to determine whether a person has balance disorder or not. The CoG data are collected in some hospitals, universities, … etc. and stored in a system made by our group. The raw data can be transformed to a new dataset which is more meaningful to doctors. We process the data by using support vector machine (SVM) technique, which is a supervised data mining technique for classification. In recent years, SVM has been applied widely in medical diagnosis. In 2013, patients with diabetes are classified by SVM with the accuracy of 78% [4]. In 2010, SVM is used for prediction of medication adherence in Heart Failure patients with 11 attributes. They tried Kernel method with 4 kernel functions and found that Radial Basis Function had the best accuracy - up to 77.6% for 2 groups of patients [5]. This paper proposed a method for processing CoG data with 2 stages. Stage 1 is to convert CoG raw data into useful parameters and to access the influence of each parameter in patient's postural balance. Stage 2 is to apply SVM for binary classification using corresponding parameters obtained in stage 1.

The paper is organized as follow. Section 2 is the methodology and the setup of the experiment. Section 3 is the results and discussion. Section 4 concludes the paper.

2 Methodology

2.1 Data Preparation and Processing

Sets of CoG data from patients with and without balance disorder are collected with CoG device. This device is made by our group to record real time CoG signals, which is shown in Fig. 1 [6].

It is a simple design with Arduino microprocessor and 4 loadcells, which is designed based on postural sway analysis. A postural sway analysis is a method based on the distribution of force in the four directions of the human body when standing on a rectangular or square board as illustrated in Fig. 2.

The device collects data on the mass that the human body produces over 4 sensors and sends signals to user interface via COM port or Bluetooth. Then, the CoG of each person is calculated as below:

$$x = \frac{[(F4 + F2) - (F1 + F3)] \times L}{W} \tag{1}$$

$$y = \frac{[(F3 + F4) - (F1 + F2)] \times L}{W} \tag{2}$$

where $W = F1 + F2 + F3 + F4$.

The platform demonstrates the projection of CoG onto 2-dimentional space, represented as a set of (x, y) coordinates. There are totally 78 participants at the age

Fig. 1. Self-made CoG device

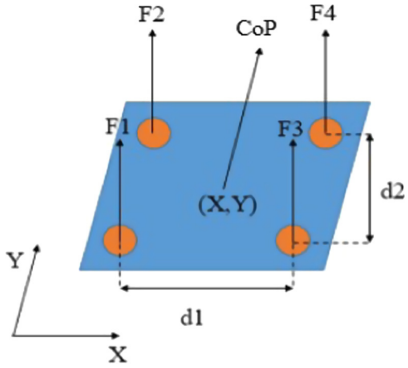

Fig. 2. CoG Platform design

between 22 and 60. All patients are required to stand on the CoG device twice, for the duration of 45 s each, with leg-open and leg-closed status. Base on each patient's medical record and treatment, patients are divided into two groups: with and without balance disorder. The patients in group "with balance disorder" are diagnosed with a disease that has balance disorder symptom, such as vestibular disorder, high blood pressure, BPPV, cerebral circulation insufficiency, etc. People in this group currently complain about high frequency of unstability and dizziness. They stayed in the hospital for less than one week (this ensures they do not undergo so many treatment procedures). The "without balance disorder" group includes healthy patients who are not diagnosed with any diseases relating to balance disorder and rarely feel dizzy and unstable. All patients do not have any leg injury that interferes them from standing on the device and do not use stimulants before measurement.

Each dataset for each patient contains 300 samples, or 300 pairs of (x, y) measured in time series, which represents the CoG of each person in 45 s. When changing from leg-open to leg-closed position, the patient's coordinate might have insignificant numerical errors, so the first 100 samples collected in the first 15 s are eliminated and the rest are kept for processing. The mean of x and y, denoted by $\overline{x}$ and $\overline{y}$, are used as the new origin coordinate around which patients sway to keep balance. According to [7], this dataset can be transformed to a new dataset which is more meaningful to doctors: mean distance, root mean squared (RMS) distance, range, mean velocity, 95% confident circle area, 95% confident ellipse area, sway area, mean frequency, fractal dimension, total power, etc. In this paper, only some of these parameters are chosen for calculation for each dataset.

The first purpose is to figure out the influence of such parameters on human balance by examining two independent groups, which is done by Independent Sample T-test. Group 1 includes 28 patients with balance disorder, group 2 includes 50 patients without balance disorder, and each patient in each group corresponds to a set of values, including mean distance, mean distance x, mean distance y, RMS distance, RMS distance x, RMS distance y, mean velocity, mean velocity x, mean velocity y, mean frequency, mean frequency x, mean frequency y, 95% confidence circle area, 95% confidence ellipse area and sway area, drawn out from the raw set of (x, y) data. When considering each parameter, the Independent Sample T-test is used to check whether the mean values of two group for each parameter are statistically different from each other. If yes, then that parameter could have impact on human's balance disorder.

The second purpose is to use SVM to classify patients with and without balance disorder, which requires above parameters as the input of SVM. The inputs of SVM should include attributes that have significant impact on human balance. According to independent sample T-test value, if there is any parameter that makes the mean values of two group not different from each other, it will be ignored temporarily, and the rest will be considered as the input of SVM. A user interface by C# programming language and Microsoft Visual Studio is created to support parameter calculations for the purpose of binary classification.

All calculations will be done in reference to the mean CoG, which means each pair (x, y) is transformed to a new pair (X, Y) in the new coordinate so that the new origin is the mean CoG ($\overline{x}$, $\overline{y}$). It is also necessary to do additional calculations for only x or y direction. The resultant distance (RD) is the distance from each data point to the mean CoG in the new coordinate.

$$X = \mathrm{x} - \overline{x}, Y = \mathrm{y} - \overline{y} \tag{3}$$

$$RD[n] = \left(X[n]^2 + Y[n]^2\right)^{1/2}, n = 1, \ldots, N \tag{4}$$

where N is the number of samples for each dataset. The mean distance is the average distance from the mean CoG and the RMS distance from the mean CoG is the RMS value of the RD, which is calculated as:

$$\text{Mean distance} = 1/\text{N} \sum RD[n] \tag{5}$$

$$\text{RMS distance} = \left[1/\text{N} \sum RD[n]^2\right] 1/2 \tag{6}$$

When considering only x direction, the mean and rms distance for each direction is:

$$\text{Mean distance x} = 1/\text{N} \sum |X[n]| \tag{7}$$

$$\text{RMS distance x} = \left[1/\text{N} \sum X[n]^2\right]^{1/2} \tag{8}$$

The total excursion is the total length of the CoG path, and is approximated by the sum of the distances between consecutive points on the CoG path

$$\text{Total path } = \sum_{n=1}^{N-1} \left[(X[n+1] - X[n])^2 + (Y[n+1] - Y[n])^2\right]^{1/2} \tag{9}$$

The total path when considering only x direction is:

$$\text{Total path x} = \sum_{n=1}^{N-1} |X[n+1] - X[n]| \tag{10}$$

The mean velocity (v) is the distance that the coordinates move in a time unit:

$$\text{v} = \text{Total path}/\text{T} \tag{11}$$

The increase in total path or mean velocity may suggest a poorer ability to keep balance. The 95% confidence circle is defined as the circle that contains 95% of the distance from the mean CoG, and the 95% confidence ellipse is defined similarly.

$$\text{Area CC} = \pi(\text{MD} + z_{0.5} s_{\text{RD}})^2 \tag{12}$$

$$\text{Area CE} = 2\pi \, F_{.05[2,n-2]} \left[s_{x^2} s_{y^2} - s_{xy^2}\right]^{1/2} \tag{13}$$

Where s_{RD}, s_x, s_y are the standard deviation of RD, X and Y, s_{xy} is the covariance; $z_{0.5} = 1.645$ is the z statistic at the 95% confidence interval. $F_{.05[2,\, n-2]}$ is the F statistic at 95% confidence level for a bivariate distribution with n data points. It has the value of 3.00 for a large number of samples (n > 120).

Sway area (AREA-SW) estimates the area enclosed by the CoG path per unit of time.

$$\text{Sway area } = \frac{1}{2T} \sum_{n=1}^{N-1} |X[n+1]Y[n] + + X[n]Y[n+1]| \tag{14}$$

The mean frequency is the rotational frequency, in revolutions per second or Hz, of the CoG if it had travelled the total excursions around a circle with a radius of the mean distance

$$\text{Mean f} = \frac{Total\ path}{2\pi MDT} = \frac{v}{2\pi MD} \tag{15}$$

2.2 Support Vector Machine Technique

Support Vector Machine Methodology. Support Vector Machine is a supervised data mining technique mainly used for classification. It aims at both minimizing misclassification and maximizing the geometric margin. The most basic form of SVM is linear SVM in which all data are linearly separable, but in fact, the dataset is too complex to simply solve by linear SVM. Fortunately, SVM can work efficiently for non-linear classification by applying kernel trick, which implicitly maps the inputs into a new space with higher dimensions to minimize nonlinear complexity. In this paper, we use a mapping function $x \rightarrow \varphi(x)$ to cast the original input data into a higher dimension space to deal with nonlinearity. However, calculating $\varphi(x)$ for each sample is very complex, especially when the dimension of the data set increases, and the number of samples is large. We need to simplify the process by using kernel trick in which only inner products (dot product) of the mapped inputs in the feature space need to be determined without explicitly calculate φ. Four kernel functions are summarized in Table 1 [8]:

Table 1. Four kernel functions

Name of kernel function	Definition
Polynomial	$k(x, z) = (r + \gamma x^T z)^d$
Laplacian	$k(x, z) = \exp(-\|\|x - y\|\|/(2\delta^2)), \gamma > 0$
Radial basis function	$k(x, z) = \exp(-\gamma\|\|x - y\|\|^2), \gamma > 0$
Sigmoid	$k(x, z) = \tanh(\gamma x^T z + r)$

The kernel trick converts the objective function into a new form:

$$\begin{aligned} \lambda &= \arg \max_{\lambda} \sum_{n=1}^{N} \lambda_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \lambda_n \lambda_m y_n y_m k(x_n, x_m) \\ \text{subject to} \quad & \sum\nolimits_{n=1}^{N} \lambda_n y_n = 0; 0 \le \lambda_n \le C, \forall n = 1, 2, \ldots, N \end{aligned} \tag{16}$$

where N is the number of data point in the training set; xn is the nth vector in the training set; yn is the label of nth data point (yn can be 1 or −1); λn is the Lagrange factor of nth data point; C is a constant described above. After this function is solved, support vectors will be found, and labelling can be performed next.

Experiment Setup. We divided the data into training and testing dataset. Training dataset includes 50 samples, in which 18 samples are labelled as 1 (which means they are in group "with balance disorder"), and others are labelled as −1 (which means they are in group "without balance disorder"). Testing dataset includes 28 samples, in which 10 samples are labeled as 1 and 18 samples are labeled as −1. The division of training and testing data is based on hold-out technique (training set/testing set = 64/36) [9].

We selected parameters using SVM library in C#. It allows parameter optimization by grid search tool, in which cross-validation accuracy is obtained for each parameter and the one with highest cross-validation accuracy will be considered the optimal value [10].

In order to evaluate our methodology, we calculate the accuracy, sensitivity and specificity for each classification. Each classification gives us values of TP (true positive), TN (true negative), FP (False positive), FN (False negative) which help to measure accuracy, sensitivity and specificity [11]. These values are critical in assessment of a specific test's reliability. Sensitivity shows the probability a test can correctly give a positive result for people who have disease. For a test with high sensitivity, it will generate positive result for almost everyone who has the disease and return just few false-negative results. Specificity of a test is the ability to correctly generate a *negative* result for people who *do not* have that disease. A test with high specificity will correctly identify almost people who *do not* have the disease and the rate for false-positive results is very low. The calculation of accuracy, sensitivity and specificity is given as below.

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \tag{17}$$

$$\text{Sensitivity (TP rate)} = \frac{TP}{TP + FN} \tag{18}$$

$$\text{Specificity (FP rate)} = \frac{FP}{FP + TN} \tag{19}$$

3 Results and Discussion

According to Independent Sample t-test, the mean values of two group for each parameter are statistically different from each other, so all above parameters have impact on human's balance and could be useful for clinical assessment. Especially, parameters calculated for solely x or y direction also reflect a significant difference between 2 groups, which is identical to the results obtained in [12]. Theoretically, patients with balance disorder will agitate more than normal people, so their coordinate also swing more around the origin. Hence, there is an increase in all above parameters, which is identical to the results of Independent Sample t-test. The p-value for each parameter is summarized in Table 2. In this t-test, the null hypothesis H0 and H1 is expressed as "the two means of the two groups are equal" and "the two means of the

two groups are not equal", respectively. When p-value is less than α ($\alpha = 0.05$), it indicates a significant difference between 2 groups and vice versa.

Table 2. p-values of all parameters when comparing means between 2 groups (group 1: n = 28; group 2: n = 50)

Parameter	Mean distance	Mean distance x	Mean distance y	RMS distance	RMS distance x	RMS distance y
P-value	0.003	0.024	0.001	0.002	0.009	0.001
Parameter	Mean velocity	Mean velocity x	Mean velocity y	Mean frequency	Mean frequency x	Mean frequency y
P-value	0.000	0.000	0.000	0.000	0.000	0.005
Parameter	95% confidence circle area		95% confidence ellipse area		Sway area	
P-value	0.002		0.001		0.000	

All parameters are eligible for the input of SVM. The first trial for SVM involves the use of 7 attributes: mean distance, RMS distance, mean velocity, mean frequency, 95% confidence circle area, 95% confidence ellipse area and sway area, in which movement of the coordinate in solely x or y direction is not taken into account, and the second one has 15 attributes, which means 8 attributes is added: mean distance x, mean distance y, RMS distance x, RMS distance y, mean velocity x, mean velocity y, mean frequency x, mean frequency y.

These two trials help to figure out whether a better result can be obtained when adding more parameters relating to the movement of patients' CoG data in one direction. When using SVM with 7 attributes, the accuracy for the training and testing data sets is obtained and described in Table 3:

Table 3. Performance of 7-attribute SVM

Kernel functions	Data set	Accuracy	Sensitivity	Specificity
Gaussian (RBF)	Training set	88%	77.2%	96.9%
	Testing test	78.6%	70%	83.3%
Polynomial	Training set	62%	100%	40.6%
	Testing set	57.1%	100%	33.3%
Laplacian	Training set	76%	33.3%	100%
	Testing set	67.9%	20%	94.4%
Sigmoid	Training set	64%	0%	100%
	Testing set	64.3%	0%	100%

When using SVM with 15 attributes, the accuracy for training and testing data sets is obtained and described in Table 4:

Table 4. Performance of 15-attribute SVM

Kernel functions	Data set	Accuracy	Sensitivity	Specificity
Gaussian (RBF)	Training set	90%	72.2%	100%
	Testing test	82.1%	80%	83.3%
Polynomial	Training set	84%	100%	75%
	Testing set	82.1%	100%	72.2%
Laplacian	Training set	90%	83.3%	93.8%
	Testing set	75%	80%	72.2%
Sigmoid	Training set	64%	0%	100%
	Testing set	64.3%	0%	100%

Referring to Table 3 and Table 4, the RBF is the most effective function among 4 kernel functions with the highest accuracy. The accuracy of RBF when performing classification on training dataset is 88% for 7 attributes and 90% for 15 attributes. The accuracy of RBF when performing classification on testing dataset is 78.6% for 7 attributes and 82.1% for 15 attributes. Sigmoid function brings up the lowest accuracy ($\sim$64%) and sensitivity (0%) so it is not a promising function. In 15-attribute SVM with 3 kernel functions, the accuracy, sensitivity, and specificity for classification of training and testing set are greater than or equal to those in 7-attribute SVM. Therefore, SVM with more attributes has better outcomes, and the movement of CoG data in solely one direction, indicated by 8 additional parameters, can influence on classification accuracy. Laplacian function and Polynomial function do not seem to work well with 7-attribute SVM but they have better results in 15-attribute SVM. We can see that through experiment, RBF is the most effective function. Moreover, SVM with 15 attributes has better results than SVM with 7 attributes since it almost increases the accuracy, sensitivity, and specificity of classification for both training and testing set.

4 Conclusion

In this paper, SVM on the CoG data is used to diagnose diseases with balance disorder. Parameters that can be used to assess human's postural balance include: mean distance, mean distance x, mean distance y, rms distance, rms distance x, rms distance y, mean velocity, mean velocity x, mean velocity y, mean frequency, mean frequency x, mean frequency y, 95% confidence circle area, 95% confidence ellipse area and sway area. When classifying patients with and without balance disorder by SVM algorithm, RBF is the function that is the most effective. Moreover, SVM with 15 attributes has better results than SVM with 7 attributes since it almost increases the accuracy, sensitivity and specificity of classification for both training and testing set. One limitation of this research is that the impact of age on human's postural balance is ignored. The CoG pattern for young and old people may differ, because the older usually has poorer ability to keep balance due to the degradation of skeleton system. This fact opens a

future direction for the topic. In conclusion, 15 parameters listed above can demonstrate patient's balance status, and SVM algorithm is a potential technique to classify patients with and without balance disorder with high accuracy.

References

1. Watson, M.A., Owen Black, F., Crowson, M.: The human balance system: a complex coordination of central and peripheral systems by vestibular disorders association. https://vestibular.org/sites/default/files/page_files/Documents/Human%20Balance%20System_36.pdf
2. Lava, N.: A visual guide to balance disorder, 10 April 2008. http://www.webmd.com
3. Robinson, B.S.: Common vestibular function tests. http://www.neuropt.org
4. Chitra, R., Anuja Kumari, V.: Classification of diabetes diseases using support vector machine. Int. J. Eng. Res. Appl. **3**(2), 1797–1801 (2013)
5. Lee, S.-K., et al.: Application of support vector machine for prediction of medication adherence in heart failure patients. Healthc. Inform. Res. **16**, 253–259 (2010)
6. Huy, H.Q., et al.: A design of a vestibular disorder evaluation system. In: Solanki, V.K., Hoang, M.K., Lu, Z., Pattnaik, P.K. (eds.) Intelligent Computing in Engineering. AISC, vol. 1125, pp. 1105–1117. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-2780-7_114
7. Prieto, T.E.: Measures of postural steadiness: differences between healthy young and elderly adults. IEEE Trans. Biomed. Eng. **43**(9), 956–966 (1996)
8. Souza, C.R.: Kernel functions for machine learning applications, 17 March 2010. http://crsouza.blogspot.com/2010/03/kernel-functions-for-machine-learning.html
9. Allibhai, E.: Hold-out vs. cross-validation in machine learning, 03 October 2018. https://medium.com
10. Hsu, C.-W., Chang, C.-C., Lin, C.-J.: A practical guide to support vector classification (2016)
11. Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., Thomas, R.: Understanding and using sensitivity, specificity and predictive values. Indian J. Ophthalmol. **56**, 45 (2008)
12. Hossein, T., et al.: Static balance in patients with vestibular impairments: a preliminary study. Scientifica **56**(1), 1–5 (2016)