Expert Systems with Applications 38 (2011) 7359–7363

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Introduction and application of an automatic gait recognition method to diagnose movement disorders that arose of similar causes

Masood Banaie [a,*], Mohammad Pooyan [b], Mohammad Mikaili [b]

[a] Department of Biomedical Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran
[b] Department of Biomedical Engineering, Faculty of Engineering, Shahed University, Tehran, Iran

ARTICLE INFO

Keywords:
Huntington's disease (HD)
Parkinson's disease (PD)
Amyotrophic Lateral Sclerosis (ALS)
Classification
Feature extraction
Gait

ABSTRACT

Nowadays, researchers try to introduce a more convenient and fast approach (especially a non-invasive one) to diagnose diseases fast and more precisely. Some patient death may be because of wrong diagnosis. Some patients that suffer diseases such as Parkinson disease are known to be dead of wrong diagnosis. Such an approach would help physicians to focus on the correct disease and its treatment and to avoid wasting precious time – that may be critical for the patient – on diagnosis. In this study, we try to develop a new automated approach for classifying (diagnosing) locomotive patients using features that may be extracted from their gait signal. We selected four groups: patients with Huntington's disease, Parkinson's disease and Amyotrophic Lateral Sclerosis and a group of healthy subjects. Examining different available classifiers on all proposed features, we have introduced a novel feature with acceptably low error rate using quadratic Bayes classifier.

1. Introduction

It is obvious that diagnosing different disorders in a correct and fast way is highly important in prescribing the correct treatment for the disease. It would be valuable if it was done by a non-invasive approach. Moreover, identifying the disease in a more simplified way will help physicians in decreasing the cost and time of the diagnosis process. Diagnosis time is an important factor in the treatment especially for progressive diseases. In such a situation that we may need medical imaging or genetic tests to detect the disease, diagnosing movement diseases that have a similar cause through their effect on the movements of the patient will be precious. Gait signal may be a good factor for discriminating movement disorders that is caused by malfunctioning of some brain parts. It can also be used for validation of models that are introduced for the diseases (Haeri, Sarbaz, & Gharibzadeh, 2005). It also can be used as a base for training models that may be introduce to model similar disease in a correct way.

In this study, we evaluate different classifiers on the features that are extracted from gait signal of healthy persons and patients with three different movement disorders including Huntington's disease (HD), Parkinson's disease (PD), and Amyotrophic Lateral Sclerosis (ALS). These diseases are selected as they have similar causes and all are neurodegenerative diseases. These disease were selected because, they are known to have similar causes that may complicate diagnosis of them. We present a brief explanation for each of them in the following lines.

HD is a genetic neurological disorder characterized by abnormal body movements called chorea and a lack of coordination. Because some of its symptoms are similar to other neurological disorders, such as Alzheimer's disease, it has been extensively studied by different researchers. If the symptoms are noticeable before a person is twenty, then their condition is called Juvenile HD. As there is currently no proven cure, symptoms are controlled with various medications and supportive services. Complications caused by its symptoms reduce life expectancy (Banaie, Sarbaz, Gharibzadeh, & Towhidkhah, 2008; Kandel, Schwartz, & Jessell, 2000).

PD is also a degenerative disorder of the central nervous system that often impairs the sufferer's motor skills and speech, as well as other functions. It is characterized by muscle rigidity, tremor, a slowing of physical movement (bradykinesia) and, in extreme cases, a loss of physical movement (akinesia). The primary symptoms are due to decreased stimulation of the motor cortex by the basal ganglia. PD is both chronic and progressive (Haeri et al., 2005; Kandel et al., 2000).

ALS is a progressive, usually fatal, neurodegenerative disease caused by the degeneration of motor neurons, the nerve cells in the central nervous system that control voluntary muscle movement. As a motor neuron disease, the disorder causes muscle weakness and atrophy throughout the body as both the upper and lower motor neurons degenerate, ceasing to send messages to muscles (Kandel et al., 2000).

* Corresponding author. Tel.: +98 935 2128208; fax: +98 21 51212021.
E-mail address: Banaie@Gmail.com (M. Banaie).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.12.091

2. Materials and methods

In this study, we have used a database from Physionet. The database includes data recorded from 15 patients suffering PD, 20 of HD, 13 of ALS and 16 healthy (control) subjects. The raw data were obtained using force-sensitive resistors, with the output roughly proportional to the force under the foot. Each record included two signals recorded from each foot of the subject (physionet, XXXX).

Recordings were first analyzed visually and we removed two records (one of PD and one of HD records); because they were erroneously recorded in a limited period of time that may be caused by sensor malfunctioning. Before preparing our test and training datasets, we removed linear trends of the data to eliminate their offset. Then, we tried different features that seemed to be useful on classification of the data. First, we tried time-domain features such as zero-crossing, mean, and standard deviation of stride-to-stride time. We used a toolbox written for MATLAB named PRTools. This toolbox consists of a complete and useful set of functions for pattern recognition and includes most of the well known classification algorithms (Duin et al., 2004). We converted the mentioned dataset to PRTools standard format. We selected a set of 17 well known classifiers that their computation procedures would be described briefly in following lines.

2.1. Linear and quadratic Bayes normal classifier

In this classifier, the pattern-generating mechanism is represented in a probabilistic framework. A Bayes classifier is a pattern classifier based on two prerequisites: (1) the damage or loss of value, involved when an object is incorrectly classified can be quantified as a cost. (2) The expectation of the cost is acceptable as an optimization criterion (Duda, Hart, & Stork, 2001; Heijden, Duin, Ridder, & Tax, 2004).

2.2. Optimized Parzen classifier

This classifier computes the optimum smoothing parameter for the Parzen classifier between the classes in the dataset. The leave-one-out Lissack and Fu estimate (Lissack & Fu, 1976) calculates the classification error. The final classifier is stored. When smoothing is specified, no learning is performed and just the discriminant is produced for the given smoothing parameters (Heijden et al., 2004).

2.3. Back-propagation trained feed-forward neural network classifier

A feed-forward neural network classifier with two hidden units in the second layer is used for the dataset. Training is stopped after at least 50 epochs or when the iteration number exceeds twice of the best classification result (Heijden et al., 2004).

2.4. K-nearest neighbor classifier

It has an estimator with high resolution in regions where the training set is dense. Therefore, the balance between resolution and variance can be adjusted locally (Heijden et al., 2004).

2.5. Linear classifier built on the Karhunen Loeve expansion of the common covariance matrix

It finds a linear discriminant function for the dataset. This is done by computing the LDC on the data projected on the first eigenvectors of the averaged covariance matrix of the classes. Either first N eigenvectors are used or the number of eigenvectors is determined such that ALF, the percentage of the total variance is explained.

2.6. Logistic linear classifier

This classifier computes a linear classification for the dataset by maximizing the likelihood criterion by means of the logistic (sigmoid) function Webb, 2002.

2.7. Fisher's least square linear classifier

It finds the linear discriminant function between the classes in the dataset by minimizing the errors in the least square sense. This is a multi-class implementation which uses the one-against-all strategy (Duda, Hart, & Stork 2001).

2.8. Nearest mean classifier (NMC) and nearest mean scaled classifier (NMSC)

It computes the linear discriminant for the classes in the dataset, assuming normal distributions with zero covariances and equal class variances. The difference with NMC is that NMSC is based on an assumption of normal distributions and thereby automatically scales the features and is sensitive to class priors. This is a plain nearest mean classifier that is feature scaling sensitive and insensitive to class priors (Veenman & Reinders, 2005).

2.9. Linear perceptron classifier

This algorithm fixes all errors encountered in the training set. In a two-class problem, the decision function is equivalent to a test $g1(y) - g2(y) > 0$. If the test fails, it is decided for 2, otherwise for 1. The test can be accomplished equally well with a single linear function:

$$g(y) = w^{\mathrm{T}} y \tag{1}$$

defined as $g(y) = g1(y) - g2(y)$. The so-called perceptron is a computational structure that implements $g(y)$. The two possible classes are encoded in the output as '1' and '−1'. In perceptron learning, the iterative procedure is stopped when all samples in the training set are classified correctly (Alpaydin & Jordan, 1996; Heijden et al., 2004).

2.10. Polynomial classifier

This classifier is also known as higher-order neural network (HONN). HONN is a single-layer network with the polynomial terms of pattern features as inputs. To overcome the exponential explosion of input units for high-dimensional data, the number of input units can be decreased by polynomial term selection, local expansion, or dimensionality reduction before polynomial expansion (Liu & Sako, 2006).

2.11. Uncorrelated normal based quadratic Bayes classifier

It is similar to quadratic Bayes classifier. But it computes a quadratic classifier between the classes in the dataset assuming normal densities with uncorrelated features (Duda et al., 2001).

2.12. Mixture of Gaussian classifier

For each class j in the dataset, a density estimate is made using two mixture components. Using the class prior probabilities, they are combined into a single classifier. The number of mixture components (two) is applied to each class.

2.13. Decision tree classifier

This method break up a complex decision into a union of several simpler decisions, hoping the final solution obtained would resemble the intended desired solution. Decision tree classification generates the output as a binary tree-like structure. A decision tree classifier is constructed in two phases: growth phase and prune phase. After building the initial tree (the 'growth phase'), a sub-tree is built with the least estimated error rate (the 'prune phase') Argentiero, Chin, & Beaudet, 1982.

2.14. Naive Bayes classifier

The Naive Bayes classifier estimates for every class and every feature separately. Total class densities are constructed by assuming independency and consequently multiplying the separate feature densities. The used default version divides each axis into 10 bins, counts the number of training examples for each of the classes in each of the bins, and classifies the object to the class that gives maximum posterior probability. Missing values will be put into a separate bin.

2.15. Levenberg–Marquardt trained feed-forward neural net classifier

A feed-forward neural network classifier with length (N) hidden layers with $N(I)$ units in layer I is computed for the dataset. Training is stopped after if the iteration number exceeds twice that of the best classification result. This is measured by the labels of the dataset. Standard Matlab initialization is used for initial weights.

These data was applied to the described classifiers and unacceptably high error rates were achieved (more than 50% in most cases). Using other features such as coefficients of a second order autoregressive with exogenous input (ARX) model and fractal dimension of the signal did not make the results much better.

Looking for a more efficient feature, we have focused on a set of frequency-domain features (i.e. band powers). There are a couple of methods to calculate band powers including FFT and DCT. FFT (fast fourier transform) was taken from the signal and its power in specific frequency bands were calculated as features. The error for this trial was similar to using the ARX coefficients. Even using a mixture of time features and frequency domain features extracted from FFT did not make the classification results reasonably better.

Finally, we have proposed our novel method on extracting features using DCT (discrete cosine transform). Because the gait signal is not stationary, we thought that DCT may result in better error rates. So, we used DCT instead of FFT to extract some features, calculating the trapped zone of DCT curve in special bands (to decrease the number of features and the computation time). We have taken DCT of all the signals and compared them visually. The formula for calculating discrete cosine transform of a signal is:

$$y(k) = w(k)\sum_{n=1}^{N} x(n)\cos\frac{\pi(2n-1)(k-1)}{2N}, \quad k = 1,\ldots,N,$$

where

$$w(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 1 \\ \sqrt{\frac{2}{N}}, & 2 \leqslant k \leqslant N \end{cases}$$

N is the length of x, and x and y are the same size. The time series are indexed from $n = 1$ and $k = 1$ instead of the usual $n = 0$ and $k = 0$ because MATLAB® vectors run from 1 to N instead of 0 to $N - 1$.

The DCT plots of the four classes are shown on Fig. 1. Defining special bands as shown in Table 1, we proposed that the signals could be fairly classified by calculating the trapped zone under the curve of each band as a feature. Despite defining such a procedure to extract the features, the features have not been classifiable easily at first look. Fig. 2 shows some scatter plots for selected features that shows the complexity of classification.

We divided the dataset to train and test sets randomly. However, we have tried to select our test set among those recordings which had been recognized as the most severe cases by a physician. So we have selected 23 of 124 recordings as test and remaining as training set. Afterwards, we trained 18 most popular classifiers that were defined on PRTools. Considering this set of features made the error rates acceptably low for some of classifiers-especially for quadratic Bayes classifier (second in the Table 2).

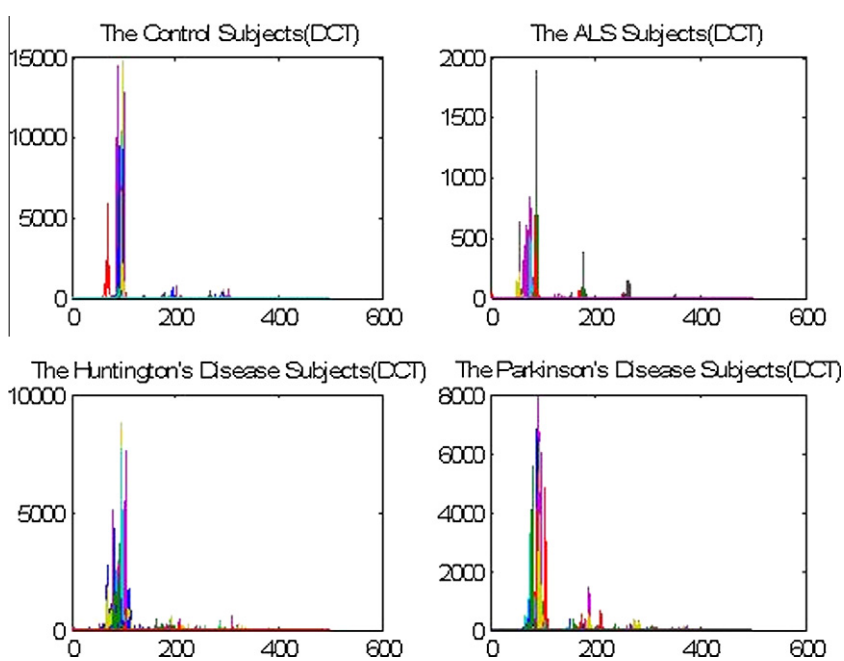

Fig. 1. DCT plots of the four classes.

Table 1
Sample bands for calculating trapped zone under the curve.

Band no.	1	2	3	4	5	6	7	8	9
Sample bands	1–50	60–74	75–90	93–105	110–150	165–185	186–200	210–260	300–end

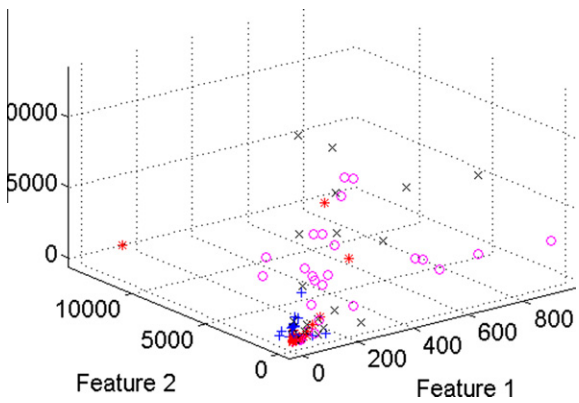

Fig. 2. Scatter plot of selected features.

The results including total error and the error that was made on diagnosing each class is shown in Table 2.

3. Discussion

An automated method for classifying movement disorders prior to further assessment and diagnosis of the severity of diseases may help physicians to focus on the correct disease and to make better decisions. So a vast majority of research work in this area is now directed toward finding automatic or semiautomatic diagnosis methods (Hiroshi, Hisanori, Takanori, & Koki, 2000; Jeon & Ko, 2007; Ross, Pritchard, Rubin, & Duse, 2006). This may result in faster and more effective treatment of the disorders. As mentioned before, this automatic classification methods may also be used introducing new models that can simply model special similar disease (that may have similar causes) in a best manner. In this study, we have tried to introduce an automated approach to diagnose three similar movement disorders through classifying the proposed features that are extracted from the gait signal of a patient. We proposed a feature set based on DCT of patient's gait signal and showed that using different classifiers, we can get reasonable error rates. According to Table 2, we conclude that using a quadratic Bayes classifier, we can classify the patients in the corresponding disease groups. Moreover, we proposed a novel feature that needs less computation for classifier to be trained and tested.

Most of studies that focus on feature extraction using DCT only use 2D-DCT and is applied to images (Kharat & Dudul, 2008; Kim & Kim, 2007; Tachaphetpiboon & Amornraksa, 2007; Yin, Fu, & Qiao, 2007). However, it is not commonly used for signals like as gait that is one dimensional. We think that DCT may improve feature extraction on non-stationary signals (such as gait signal). The trapped zone under the DCT curve in special bands is introduced to have a high capability to be used in similar studies. This approach has the advantage that uses fewer features – only the trapped zone under the curve of the nine distinct bands of the discrete cosine transform (DCT) of the gait signal – to classify the disease of the subjects in a correct manner. This can be used to prepare an online approach to diagnose the diseases in a faster and more accurate way. Moreover, it can be used training computational models of the disease beneficiating the less computation time to calculate their parameters.

4. Future considerations

In this study, we have focused on introducing a new automated approach on classifying the movement disorders in a correct manner. However, the design and implementing the classifier clinically may be considered in future.

Moreover, we had a limited access to clinical recordings needed for the train and test process. So, we could not get more accuracy. It was because the database we have used included only a limited number of subjects. Thus, training the classifier using more recordings might result in better results.

Our group is going to use this approach introducing a new universal model that would consider the states of all the disease.

Table 2
The results for different classifiers (classification accuracy of each class and the total).

Classes:	Control	Huntington	Parkinson	ALS	Total accuracy (%)
1: Linear Bayes normal classifier	100	71.429	40	50	65.217
2: Quadratic Bayes normal classifier	**100**	**71.429**	**80**	**100**	**86.957**
3: Parzen classifier	100	57.143	60	16.667	56.522
4: BP Neural classifier	80	71.429	0	83.333	60.87
5: K-NN classifier	100	42.857	60	66.667	65.217
6: KL Bayes-normal-1*	100	71.429	40	50	65.217
7: Logistic classifier	20	28.571	40	83.333	43.478
8: Fisher	80	71.429	40	50	60.87
9: Nearest mean	100	0	60	33.333	43.478
10: Scaled nearest mean	100	57.143	20	33.333	52.174
11: Linear perceptron classifier	40	42.857	60	100	60.87
12: Polynomial classifier	80	71.429	40	50	60.87
13: Uncorrelated normal based quadratic Bayes classifier	100	42.857	40	100	69.565
14: Mixture of Gaussian classifier	0	71.429	0	100	47.826
15: Decision tree	100	85.714	40	83.333	78.261
16: Naïve Bayes	100	57.143	40	33.333	56.522
17: Levenberg–Marquardt trained feed-forward neural net classifier	40	71.429	40	83.333	60.87

* Linear classifier built on the KL expansion of the common covariance matrix.

References

Alpaydin, E., & Jordan, M. I. (1996). Local linear perceptrons for classification. *IEEE Transactions on Neural Networks, 7*(3), 788–794.

Argentiero, P., Chin, R., & Beaudet, P. (1982). An automated approach to the design of decision tree classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-4*, 51–57.

Banaie, M., Sarbaz, Y., Gharibzadeh, S., & Towhidkhah, F. (2008). Huntington's disease: Modeling the gait disorder and proposing novel treatments. *Journal of Theoretical Biology, 254*(2), 361–367. 21 September 2008.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: John Wiley and Sons.

Duin, R. P. W., Juszczak, P., Paclik, P., Pekalska, E., de Ridder, D., & Tax, D. M. J. (2004). *PRTools4, A Matlab toolbox for pattern recognition*. Delft University of Technology.

Haeri, M., Sarbaz, Y., & Gharibzadeh, S. (2005). Modeling the Parkinson's tremor and its treatments. *Journal of Theoretical Biology, 236*(3), 311–322.

Heijden, F., Duin, R. P. W., Ridder, D., & Tax, D. M. J. (2004). *Classification, parameter estimation and state estimation, an engineering approach using MATLAB*. Chichester, England: John Wiley & Sons Ltd.

Hiroshi, K., Hisanori, T., Takanori, K., & Koki, K. (2000). Automatic diagnosis of mental disease by face image processing. *Nippon Kikai Gakkai Robotikusu, Mekatoronikusu Koenkai Koen Ronbunshu, 2000*(Pt.4). PAGE.2P1.12.010(1)-2P1.12.010(2).

Jeon, B., & Ko, I. (2007). Ontology-based semi-automatic construction of Bayesian network models for diagnosing diseases in E-health applications. *Frontiers in the Convergence of Bioscience and Information Technologies*.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science* (4th ed.). New York: McGraw-Hill.

Kharat, G. U., & Dudul, S.V. (2008). Neural network classifier for human emotion recognition from facial expressions using discrete cosine transform. In *First international conference on emerging trends in engineering and technology*.

Kim, H., & Kim, W. (2007). Computerized bone age assessment using DCT and LDA. In A. Gagalowicz, & W. Philips (Eds.), *MIRAGE 2007, LNCS 4418* (pp. 440–448).

Lissack, T., & Fu, K. S. (1976). Error estimation in pattern recognition via L-distance between posterior density functions. *IEEE Transactions on Information Theory, 22*, 34–45.

Liu, C. L., & Sako, H. (2006). Class-specific feature polynomial classifier for pattern classification and its application to handwritten numeral recognition. *Pattern Recognition, 39*, 669–681.

<http://www.physionet.org>.

Ross, N. E., Pritchard, C. J., Rubin, D. M., & Duse, A. G. (2006). Automated image processing method for the diagnosis and classification of malaria on thin blood smears. *Medical and Biological Engineering and Computing, 44*, 427–436. doi:10.1007/s11517-006-0044-2.

Tachaphetpiboon, & S., Amornraksa, T. (2007) Fingerprint Features extraction using curve-scanned DCT coefficients. In *Proceedings of Asia-Pacific conference on communications*.

Veenman, C. J., & Reinders, M. J. T. (2005). The nearest subclassifier: a compromise between the nearest mean and nearest neighbour. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(9), 1417–1429.

Webb, A. (2002). *Statistical Pattern Recognition*. New York: John Wiley & Sons.

Yin, H., & Fu, P., Qiao, J. (2007). Face recognition based on DCT and 2DLDA. In *Second international conference on innovative computing, information and control, 5–7 September, 2007, ICICIC'07*. (pp. 581–581).