Detection of Parkinson's Disease Through Static Analysis of Handwriting and Character Recognition

Kalana R[1], Leya Elizabeth Sunny[2]

[1] Student, Dept. Computer Science and Engineering, Mar Athanasius College of Engineering, Kothamangalam, India

[2] Professor, Dept. Computer Science and Engineering, Mar Athanasius College of Engineering, Kothamangalam, India

---***---

Abstract -Detection of changes in micrographia is a symptom in Parkinson's disease (PD). A quantitative analysis of handwriting samples would be valuable as it could supplement and support clinical assessments, help monitor micrographia, and link it to PD. Such a analysis would be especially useful if it could detect subtle yet relevant changes in handwriting morphology, thus enhancing solution of the detection procedure. It is a new method for detecting the Parkinson's disease through the static analysis of handwriting. It improves the opportunity for testing the disease, Since there are no laboratory tests like blood based tests exist to determine the existence or extent of PD. Examination of handwriting presents an enormous opportunity for tracking PD conditions over time , in and away from the clinic. The approach has potential to examine simple writing samples of patients in a natural environment. Since it only evaluating the handwriting samples for the disease prediction, this method is less costly. The accuracy of the system get increased as the number of training samples get increased. The system also recognizes the characters from the handwritten samples of PD patients.

Key Words: Parkinson's disease, Features

1. INTRODUCTION

Detection of changes in micrographia is a symptom in Parkinson's disease (PD).A quantitative analysis of handwriting samples would be valuable as it could supplement and support clinical assessments, help monitor micrographia, and link it to PD. Such a analysis would be especially useful if it could detect subtle yet relevant changes in handwriting morphology, thus enhancing solution of the detection procedure.

Currently, there are no laboratory tests exist to determine the existence or extent of PD . The efficacy of any intervention is generally determined via detection of changes in symptoms they have. The gold standard in this regard is a physical examination administered by a neurologist, who may score symptomatic severity using a scale like the Unified Parkinson's Disease Rating Scale (UPDRS). While this process is quite reliable, established, and has been modified over years of experience with clinicians including recent improvements with MDS-UPDRS , it remains relatively subjective and there is certainly potential to improve its resolution to capture symptomatic

responses to therapy, which can be fine . The requirement of a clinical visit for this also limits this process. Examination of handwriting presents an enormous opportunity for tracking PD conditions over time, in and away from the clinical infrastructure.

Handwriting is a task that is strongly affected by PD and its debilitation may be the first and important observable sign of the disease.

Parkinson's disease (PD) is a degenerative neurological disorder. The main cause of Parkinson's disease is actually unknown untill now. However, it has been researched that the combination of environmental and genetic factors play an important role in causing PD for people. For general understanding the Parkinson's disease is treated as disorder of the central nervous system which is the result of loss of cells from various parts of the brain in people.

There are mainly four cardinal features of PD that can be grouped under the acronym TRAP: Tremor at rest, Rigidity, Akinesia and Postural instability. In addition, flexed posture and freezing (motor blocks) have been included among the features of parkinsonism, with PD as the most common form. Because of the diverse profiles and lifestyles of those people affected by PD, motor and non-motor impairments should be evaluated in the context of each patient's needs and goals. Most of these scales have not been fully evaluated for validity and reliability. The Hoehn and Yahr scale is commonly used to compare groups of patients and provide gross assessment of disease progression, ranging from stage 0 (no signs of disease) to stage 5 (wheelchair bound or bedridden unless assisted). The Unified Parkinson's Disease Rating scale (UPDRS) is the most well established scale for assessing disability and impairment. Rating scales are used for the evaluation of motor impairment and disability in patients with PD.

2. SUPERVISED LEARNING

Supervised learning is the machine learning task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function,

which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a reasonable way.

In machine learning, one aims to construct algorithms that are able to learn to predict a certain target output. To achieve this, the learning algorithm is presented some training examples that demonstrate the intended relation of input and output values. Then the learner is supposed to approximate the correct output, even for examples that have not been shown problem cannot be solved exactly. The kind of necessary assumptions about the nature of the target function are subsumed in the phrase inductive bias.

Data set

The dataset used for the detection of the parkinson's disease is the handwriting samples collected from the PD patients and healthy people. Data samples from the patients and healthy people are used for the feature extraction to detect the disease.

Preprocessing

Image preprocessing is crucial in the recognition pipeline for correct disease and character prediction cases. These methods typically include noise removal, image segmentation, cropping, scaling, etc. In this project, these methods have mainly been used when recognizing from an image, but some of them, such as cropping the written character and scaling it to our input size, are also performed. This step is important in the case of disease prediction stage .Because for predicting the disease using the handwriting samples we use the differences in their writings. Therefore each of the differences in the images are important. So in order to obtain the accurate result these images should undergo preprocessing stage.

Feature Extraction

Features of input data are the measurable properties of observations, which one uses to analyze or classify these instances of data. The task of feature extraction is to choose relevant features that discriminate the instances well and are independent of each other. Selection of a feature extraction method is probably the single most important factor in achieving high recognition performance. There is a vast amount of methods for feature extraction from character images, each having different characteristics, invariance properties, and reconstructability of characters. Different features are used here for the prediction of disease from the character samples are sum of mean and standard deviation, energy, eigen values, anisotropy and orientation. After the completion of preprocessing the corresponding input image is represented as a collection of matrix elements. Mean indicate the mean of all matrix elements and standard deviation indicate the standard deviation of all matrix elements. Structure tensor is an image texture analysis

technique often used in image processing and computer vision. Structure tensor, J, of an image is a matrix derived from the image partial derivatives. It is defined as,

$$J = \begin{bmatrix} <f_x, f_x>_w & <f_x, f_y>_w \\ <f_x, f_y>_w & <f_y, f_y>_w \end{bmatrix}$$

Where fx and fy are the partial derivatives of image f(x,y) along the x and y directions respectively. It is defined for each pixel as a second order symmetric positive matrix.

The eigenvalues 1 and 2 and the corresponding eigen vectors e1 and e2 summarize the distribution of the gradient of the image within the window defined by w.

If an eigenvalue is zero, the grey values in the direction of the corresponding eigen vector do not change. If one eigenvalue is zero and one greater than zero, it represents a simple neighborhood with ideal orientation. An isotropic structure is observed when 1 = 2.

Local orientation, anisotropy and energy for each pixel can be calculated for the structure tensor matrix.

The anisotropy measure gives a relation between the length of the orientation vector to the length of the gradient vector. The values of anisotropy measure vary from 0, indicating isotropic to 1 indicating highly oriented structures.

3. SUPPORT VECTOR MACHINE

IRJET In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on..

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

When data is not labeled, a supervised learning is not possible, and an unsupervised learning is required, that would find natural clustering of the data to groups, and map new data to these formed groups. The clustering algorithm which provides an improvement to the support vector

machines is called support vector clustering (SVC) and is highly used in industrial applications either when data is not labeled or when only some data is labeled as a preprocessing for a classification pass; the clustering method was published.
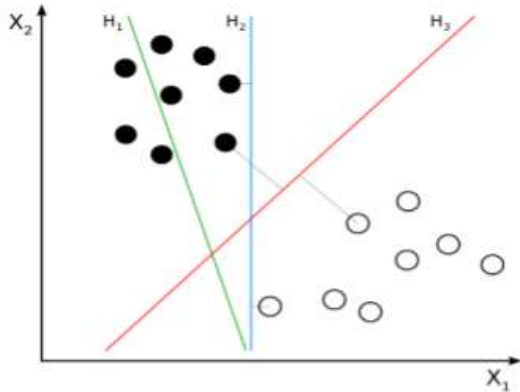

Fig-1:SVM Classification

4. PROPOSED METHOD

A. DISEASE DETECTION

A new method for detecting the Parkinson's disease through the static analysis of handwriting. It improves the opportunity for testing the disease. Currently, there are no laboratory tests like blood based tests exist to determine the existence or extent of PD. Examination of handwriting presents an enormous opportunity for tracking PD conditions over time , in and away from the clinic. The approach has potential to examine simple writing samples of patients in a natural environment. This work also provide the recognition of characters written by the PD patients which we can't read easily.

The writing samples of both the PD patients and healthy people were collected. These samples are collected in the form of images. So it undergoes some pre-processing stages to remove the unwanted data and thereby obtain the area of interest. after the preprocessing stage it undergo the feature extraction stage, to obtain the required features from the data set. Using the obtained features from both group of data set training process will be conducted. finally the classification process classify the input data set to any one of the category, I.e. patient or healthy.

In sample data collection stage data samples, ie handwriting samples were collected from both patients and healthy people. These data samples can be used for the feature extraction. The writing samples of PD patients are different from the healthy people. The characters written by the patients are not easily readable by human beings. As the number of dataset get increased the accuracy of the prediction will get more accurate.
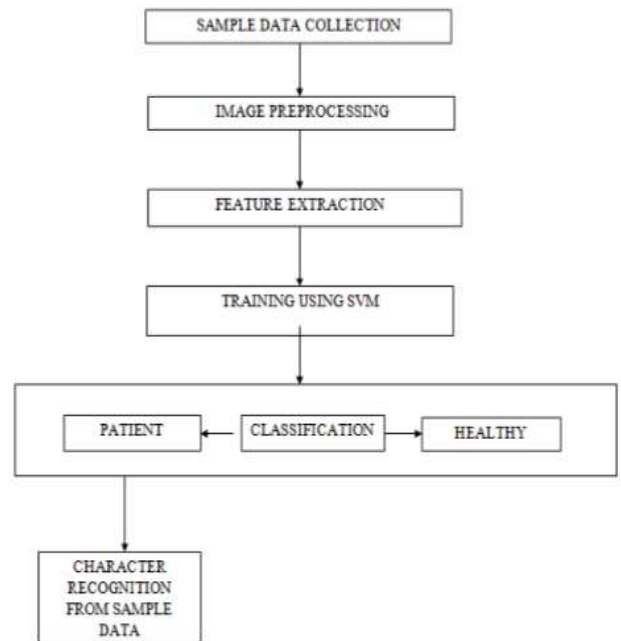

Fig-2: Proposed System Architecture

Image preprocessing is crucial in the recognition pipeline for correct disease and character prediction cases. These methods typically include noise removal, image segmentation, cropping, scaling, etc. In this project, these methods have mainly been used when recognizing from an image, but some of them, such as cropping the written character and scaling it to our input size, are also performed.

Digital capture and conversion of an input image often introduces noise which makes it hard to decide what is actually a part of the object of interest and which is not. Considering the problem of character recognition, we want to reduce as much noise as possible, while preserving the strokes of the characters, since they are important for correct classification into the correct character..

This step is important in the case of disease prediction stage also. Because for predicting the disease using the handwriting samples we use the differences in their writings. So each of the differences in the images are important. So in order to obtain the accurate result these images should undergo preprocessing stage.

Features of input data are the measurable properties of observations, which one uses to analyze or classify these instances of data. The task of feature extraction is to choose relevant features that discriminate the instances well and are independent of each other. Selection of a feature extraction method is probably the single most important factor in achieving high recognition performance. There is a vast amount of methods for feature extraction from character images, each having different characteristics, invariance properties, and reconstructability of characters.

Different features are used here for the prediction of disease from the character sample. They include energy, eigen value. anisotrpy, orientation.

Structure tensor is an image texture analysis technique often used in image processing and computer vision. Structure tensor, J, of an image is a matrix derived from the image partial derivatives. It is defined as,

$$ J = \begin{bmatrix} <f_x, f_x>_w & <f_x, f_y>_w \\ <f_x, f_y>_w & <f_y, f_y>_w \end{bmatrix} $$

The different features that are needed can be calculated by using the structure tensor. Where fx and fy are the partial derivatives of image f(x,y) along the x and y directions respectively. It is defined for each pixel as a second order symmetric positive matrix. The eigenvalues 1 and 2 and the corresponding eigen vectors e1 and e2 summarize the distribution of the gradient of the image within the window defined by w. If an eigenvalue is zero, the grey values in the direction of the corresponding eigen vector do not change. If one eigenvalue is zero and one greater than zero, it represents a simple neighborhood with ideal orientation. An isotropic structure is observed when 1 = 2.Local orientation, anisotropy and energy for each pixel can be calculated for the structure tensor matrix. The direction of the largest Eigen vector of the tensor corresponds to the local predominant orientation given by

$$ \theta = \frac{1}{2} arctan(2 \frac{<f_x, f_y>_w}{<f_y, f_y>_w - <f_x, f_x>_w}) $$

The anisotropy measure AI is given as,

$$ AI = \frac{\sqrt{(<f_y, f_y>_w - <f_x, f_x>_w)^2 + 4<f_x, f_x>_w}}{<f_x, f_x>_w + <f_y, f_y>_w} $$

The anisotropy measure gives a relation between the length of the orientation vector to the length of the gradient vector. The values of anisotropy measure vary from 0, indicating isotropic to 1 indicating highly oriented structures Energy is given by the trace of the structure tensor J,

$$ E = <f_x, f_y>_w + <f_y, f_y>_w $$

B. CHARACTER RECOGNITION

In the character recognition phase the unreadable characters from the handwriting sample of patients were recognizing. Neural network can be used for the classification of characters into it's own class.

character recognition is done by using the canny edge detection algorithm and histogram of features.The Process of Canny edge detection algorithm can be broken down to 4 different steps:

• Apply Gaussian filter to smooth the image in order to remove the noise

Find the intensity gradients of the image Apply non-maximum suppression to get rid of spurious response to edge detection

• Apply double threshold to determine potential edges

• Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

Supervised learning is the machine learning task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object, typically a vector and a desired output value .It also called as the supervisory signal. Here support vector machines(svm) are used for the training and classification.

The features selected are then sent to the classification block. Here the binary classification is done using the svm. The two classes are healthy and patient. An input image provided will be in any of these two categories.

4. CONCLUSION

A new method for detecting the Parkinson's disease through the static analysis of handwriting. It improves the opportunity for testing the disease, Since there are no laboratory tests like blood based tests exist to determine the existence or extent of PD. Examination of handwriting presents an enormous opportunity for tracking PD conditions over time , in and away from the clinic. The approach has potential to examine simple writing samples of patients in a natural environment. Since it only evaluating the handwriting samples for the disease prediction, this method is less costly. The accuracy of the system get increased as the number of training samples get increased. The character recognition for the PD patients were helps to recognize the characters written by PD patients.

ACKNOWLEDGEMENT

REFERENCES

[1] Naiqian Zhi, Beverly Kris Jaeger, Andrew Gouldstone, Rifat Sipahi, and Samuel Frank."Toward Monitoring Parkinson's through Analysis of Static Handwriting

Samples: A Quantitative Analytical Framework ".vol. 4, no.2, pp. 195-205,2015.

[2] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," Journal of Neurology, Neurosurgery and Psychiatry, vol. 79, pp. 368-376, 2008.

[3] D. Weintraub, C.L. Comella, and S. Horn, "Parkinson's disease—part 1:pathophysiology, symptoms, burden, diagnosis, and assessment," American Journal of Managed Care, vol. 14, suppl 2, pp. S40-S48, 2008.

[4] E. Tolosa, G. Wenning, and W. Poewe, "The diagnosis of Parkinson's disease," Lancet Neurology, vol. 5, pp. 75-86, 2006.

[5] A. R. Hipkiss, "Aging risk factors and Parkinson's disease: contrasting roles of common dietary constituents," Neurobiology of Aging, vol. 35, no. 6, pp.1469-1472, 2014.

[6] C.M.Toxopeus, B.M. De Jong, G. Valsan, B.A. Conway, J.H. van der Hoeven, K.L. Leenders, and N.M. Maurits, "Impairment of gradual muscle adjustment during wrist circumduction in Parkinson's Disease," PLOS ONE,vol. 6, no. 9, pp. e24572, 2011.

[7] H.R. Siebner, A. Ceballos-Baumann, H. Standhardt, C. Auer, B. Conrad,and F. Alesch, " Changes in handwriting resulting from bilateral high frequency stimulation of the sub thalamic nucleus in Parkinson's disease,"Movement Disorders, vol.14, no. 6, pp. 964-971, 1999.

[8] M. R, Lemke, H. M. Brecht, J. Koester, P.H. Kraus, and H. Reichmann,"Anhedonia, depression, and motor functioning in Parkinson's disease during treatment with pramipexole," The Journal of Neuro psychiatry Clinical Neurosciences, vol. 17, no. 2, pp.214-220, 2005.

[9] P. P´eran, D. Nemmi, D. M´eligne, A. Peppe, O. Rascol, C. Caltagirone, J.F. Demonet, and U. Sabatini, "Effect of levodopa on both verbal and motor representations of action in Parkinson's disease: A fMRI study," Brain and Language, vol. 125, pp. 324-329, 2013.

[10] R. T. Bartus, M. S. Weinberg, and R.J. Samulski, "Parkinson's disease gene therapy: success by design meets failure by efficacy," Molecular Therapy, vol. 22, no. 3, pp. 487-497, 2014..

[11] G. Ebersbach, U. Grust, A. Ebersbach, B. Wegner, F. Gandor, and A.A. Ku¨hn, "Amplitude-oriented exercise in Parkinson's disease: a randomized study comparing LSVT-BIG and a short training protocol," Journal of Neural Transmission, vol. 122, no. 2, pp. 1-4, 2014..

[12] A.I. Dumer, H. Oster, D. McCabe, L. A. Rabin, J. L. Spielman, L.O. Ramig, and J.C. Borod, "Effects of the Lee Silverman Voice Treatment(LSVT RLOUD) on

hypomimia in Parkinson's disease," Journal of the International Neuropsychological Society, vol. 20, no. 3, pp. 302-312, 2014.

[13] C. Fox, G. Ebersbach, L. Ramig, and S. Sapir, "LSVT LOUD and LSVT BIG: behavioral treatment programs for speech and body movement in Parkinson disease," Parkinson's Disease, vol. 2012, 2012.

[14] C. Fox, G. Ebersbach, L. Ramig, and S. Sapir, "LSVT LOUD and LSVT BIG: behavioral treatment programs for speech and body movement in Parkinson disease," Parkinson's Disease, vol. 2012, 2012.

[15] R.C. Zietsma, "Apparatus for use in diagnosing and/or treating neurological disorder," United States Patent. US2013/0060124 A1. Mar 7, 2013.

[16] Yinxia Liu, Weidong Zhou, Qi Yuan, and Shuangshuang Chen, "Automatic Seizure Detection Using Wavelet Transform and SVM in Long-Term Intracranial EEG", IEEE Transactions On Neural Systems And Rehabilitation Engineering,Vol. 20, No. 6, November 2012, pp 749 - 755.