Profiling Bell's Palsy based on House-Brackmann Score

Insu Song[1], Nguwi Yok Yen[1], John Vong[1], Joahchim Diederich[2], and Peter Yellowlees[3]

[1]School of Business/IT
James Cook University
Singapore Campus
insu.song@jcu.edu.au

[2]University of Melbourne
Australia
j.diederich@unimelb.edu.au

[3]Department of Psychiatry and
Behavioral Sciences, UC Davis
peter.yellowlees@ucdmc.ucdavis.edu

***Abstract*—In this study, we propose to examine facial nerve palsy using Support Vector Machines (SVMs) and Emergent Self-Organizing Map (ESOM). This research seeks to analyze facial palsy domain using facial features and grade the degree of nerve damage according to House-Brackmann score. Traditional evaluation methods involve a medical doctor taking a thorough history of a patient and determines the onset of the paralysis, the rate of progression and etc. The most important step is to assess the degree of voluntary movement present and document the grade of facial paralysis using House-Brackmann score. The significance of this work is that we attempt to apprehend this grading using semi-supervised learning with the aim of automating this grading process. The value of this research stems from the fact that there is a lack of literature seen in this area. The use of automated grading system greatly reduces assessment time and increases consistency because references of all palsy images are stored to provide references and comparison. The proposed automated diagnostics methods are computationally efficient making them ideal for remote assessment of facial palsy, profiling of a large number of facial images captured using mobile phones and digital cameras.**

Keywords—Facial, palsy, SVM, face, SOM, Health Informatics, eHealth, Medical Data Analysis.

I. INTRODUCTION

Bell's palsy or idiopathic facial paralysis (IFP) is a frequent disease worldwide, affecting all ages, without gender distinction and with high rate of remission [1]. It has been reported as an adverse event following immunization (AEFI). There were some confusion over the definition of Bell's palsy. According to Webster's Medical Dictionary [2], Bell's palsy is the paralysis of facial nerve, the nerve that supplies the facial muscles on one side of the face. The facial nerve controls the muscles of facial expression, paralysis, which results in a lack of facial expression. This is not only an aesthetic issue, but also has functional consequences as the patient cannot communicate effectively [3]. The cause of facial nerve paralysis is thought to be due to viruses. The disease typically starts suddenly and causes paralysis of muscles of one side of the face, on which the facial nerves are affected [2]. Patients often complain of a peculiar sensation of the face on the side of paralysis. This is usually attributable to faulty feedbacks from the paralyzed muscles [4]. About 80% of patients recover within weeks to months. The diagnosis of Bell's palsy is important in order to provide timely and effective treatments.

Diagnosis of IFP poses greater challenge in the case of facial paralysis in children. It is more difficult to perform an electrodiagnostic test of facial function in infants or young children than in adults because children are often not cooperative and has limited ability to assist [5]. This makes an automated diagnosis tool highly desirable. Kasse et. al. [1] reviewed 1,521 clinical cases of Bell's palsy and their prevalence according to sex, age, paralyzed side, installation mode, previous symptoms, associated symptoms, and minimal electrical stimulation tests. They concluded that clinical factors and minimal electrical simulation tests could be used to indicate the prognosis of facial palsy when electroneurography is not available.

The standard method of measuring facial nerve functions is using the House-Brackmann facial nerve grading system, which was introduced in 1983 [6], and endorsed by the Facial Nerve Disorders Committee of the American Academy of Otolaryngology in 1984. This grading scale has been thought to accurately describe a patient's facial function and to monitor their status over time to assess the course of recovery and the effects of treatment. It is also a scale that can be used rapidly enough to be useful in clinical practices [7]. However, it was developed as a gross scale, with the objective of placing patients in general categories. Therefore, it has wide application and has been found to be reliable. Table 1 shows the House-Brackmann Grading System (HBGS). It has been noted that the HBGS base around functioning of eyes closure, mouth and forehead movement to determine estimated function. The House-Brackmann Score further grade the degree of nerve damage in facial nerve palsy from 0% (Grade VI) to 100% (Grade I). We propose to grade Bell's palsy based on the HBGS using Support Vector Machines (SVMs) [8] and ESOM [9] which are described in Section II.

This automated assessment method of Bell's palsy can be implemented on smart phones, such as Android mobile devices. The proliferation of mobile phones now enables us to collect high quality image data conveniently at home, even by parents of children. In particular, digital cameras of cell phones provide easy to use interfaces for capturing useful information about general well beings of individuals. Mobile devices have been used in rapid diagnosis of health conditions [10], health

978-1-4673-5883-5/13/$31.00 ©2013 IEEE

Table 1. House-Brackmann Grading System

Face	Grade	Characteristics
Forehead	I. Normal	Normal function
	II Mild Dysfunction	Slight weakness to good function
	III. Moderate Dysfunction	Noticeable slight to moderate movement
	IV. Moderately Severe Dysfunction	Obvious weakness or disfiguring asymmetry
	V. Severe Dysfunction	Barely perceptible motion
	VI. Total Paralysis	No movement
Eye	I. Normal	Normal function
	II Mild Dysfunction	Complete closure with minimal effort
	III. Moderate Dysfunction	Obvious weakness, eye closure with effort
	IV. Moderately Severe Dysfunction	Incomplete eye closure
	V. Severe Dysfunction	Barely perceptible eyelid movement
	VI. Total Paralysis	No movement
Mouth	I. Normal	Normal function
	II Mild Dysfunction	Slight asymmetry or weakness of mouth movement
	III. Moderate Dysfunction	Obvious but no disfiguring weakness
	IV. Moderately Severe Dysfunction	Asymmetry at rest
	V. Severe Dysfunction	Barely perceptible mouth movement
	VI. Total Paralysis	No movement

social networks [11], and improving livelihood of rural areas of developing countries [12]. This rapid penetration of mobile devices worldwide now opens up new opportunities to provide more convenient and cost effective means of diagnosing various health conditions associated with Bell's palsy, such as cold and flu.

II. BACKGROUND

A. Support Vector Machine

Support Vector Machine was originally a linear classifier based on optimal hyperplane algorithm developed by Vapnik in 1963 [13]. In 1992, Bernhard Boser, Isabelle Guyon and Vapnik successfully apply kernel method to a maximum-margin hyperplane and build a non-linear classifier [14]. In 1995, Cortes and Vapnik [15] suggested a modified maximum margin classifier called soft margin classifier that allows misclassified data with a soft margin. If there is no hyperplane that can separate the data into two classes, the soft margin classifier selects a hyperplane that separates the data as cleanly as possible with maximum margin.

Burges [16] gives a tutorial on Support Vector Machines and its application on pattern recognition. Joachims [17] in 1997 published that SVM outperformed all other major classifiers in text classification. Although SVM is naturally binary-classifier, Rennie and Rifkin [18] have applied SVM in multi-class text classification and compared with naive Bayesian classifier. In their experiment, SVM gave much lower error scores than Naive Bayes.

SVMs are based on the structural risk minimization principle in order to overcome the overfitting problems. Support vector machines find the hypotheses out of the hypothesis space H of a learning system which approximately minimizes the bound on the actual error by controlling the empirical error using training samples and the complexity of the model using the VC-dimension of H. SVMs are very universal learning systems [17]. In their basic form, SVMs learn maximal margin hyperplanes (linear threshold functions). A hyperplane can be defined by a weight vector w and a bias b:

$$w \cdot x + b = 0$$

The corresponding threshold function for an input vector x is then given by:

$$f(x) = \mathrm{sign}(w \cdot x + b)$$

However, it is possible learn polynomial classifiers, radial basis function (RBF) networks and three or more layered neural networks by mapping input data x to some other (possibly infinite dimensional) feature space $\varphi(x)$ and using kernel functions $K(x_i, x_j)$ to obtain dot products, $\varphi(x_i) \cdot \varphi(x_j)$, of feature data.

B. Emergent Self-Organizing Map

Emergent Self-Organizing Map (ESOM) [19], also referred to as a Kohonen map having a large number of map neurons, is a technique using nonlinear projection neurons arranged in a map M, usually in 2D (i.e., $M \subset \Re^2$), for low-dimensional visualization of high dimensional data. Two types of ESOM grid structures have been commonly in use: Hexgrid (honeycomb like) and Quadgrid (as a lattice) maps.

ESOM usually forms a (2D) grid of prototype neurons, which are represented using high-dimensional vectors having the same size as the number of features of input data. The density of data in the vicinity of the neurons, and the distances

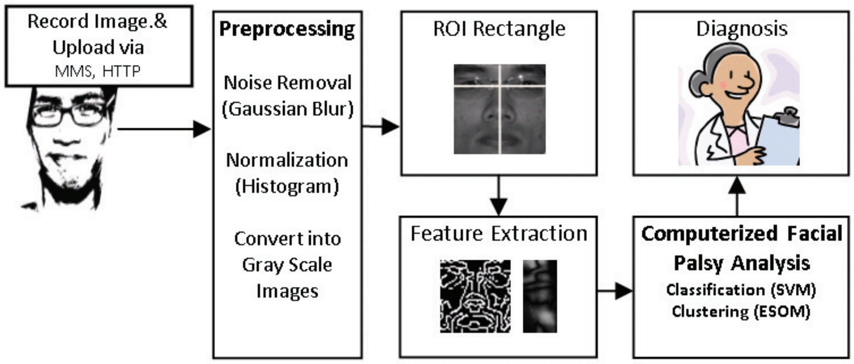


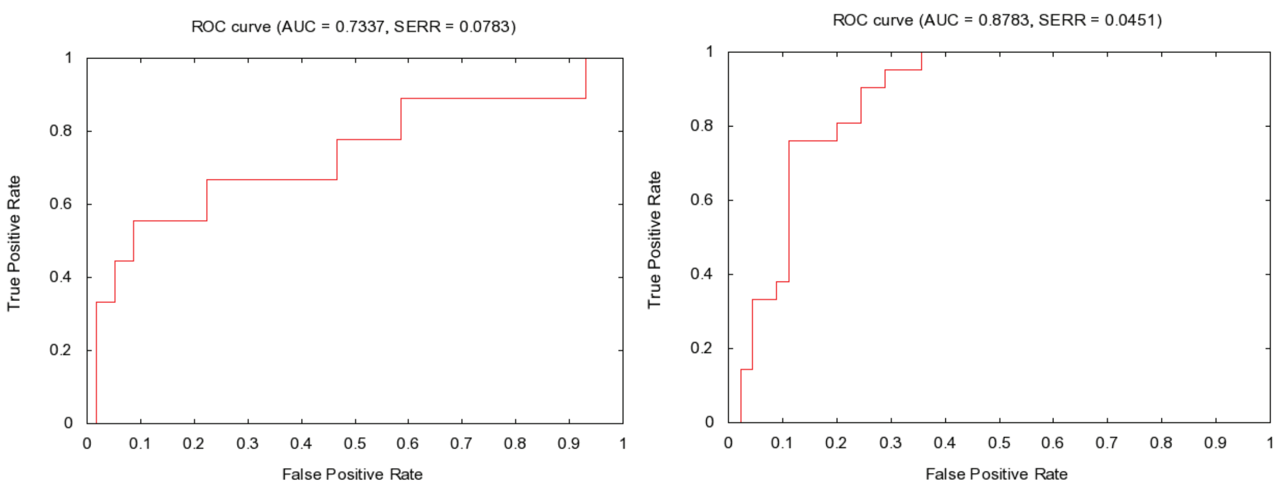
Figure 1 The overall process of automated House-Brackmann Grading of face images.

(a) Original (b) Edge (c) Difference

Figure 2 Intermediate Images of normal face. (a) Original face. (b) Significant face edges (c) Difference between left and right face.

(a) Original (b) Edge (c) Difference

Figure 3 Intermediate Images of palsy face (Grace V). (a) Original face. (b) Significant face edges (c) Difference between left and right face.



(a) (b)

Figure 4 ROC using (a) Ten principal components (b) Hamming Distances of left face and right face

between the neurons, are taken into account for better visualization. For example, an ESOM map can be visualized using

1. Distance-based Visualization called U-Map (from U-Matrix) [20];
2. Density-based Visualization called P-Map (from P-Matrix) [21]; or
3. Distance- and Density-based Visualization called U*-Map (which combines the U and P map) [22].

These different types of maps visualize the high dimensional data space on a floor space layout using a landscape like visualization for distance and/or density structure of the high dimensional data space. Structures emerge on top of the map by the cooperation of many neurons during the learning phase. These emerging structures are the main concept of ESOM. It can be used to achieve visualization, clustering, and classification. This can be obtained by the following.

Let $m{:}D \rightarrow M$ be a mapping from a high dimensional data space $D \subset \mathfrak{R}^n$ onto a finite set of positions $M=\{n_1,...,n_k\} \subset \mathfrak{R}^2$ arranged on a grid. Each position has its two dimensional coordinates and a weight vector $\boldsymbol{w}=\{w_1,...,w_k\}$, which is the image of a Voronoi region in D: the data set $E=\{x_1,...,x_d\}$ with $x_i \in D$ is mapped to a position in M such that a data point x_i is mapped to its best-match $b_m(x_i) = n_b \in M$ with $d(x,w_b) \leq d(x,w_j)$; $\forall w_j \in \mathbf{W}$, where d is the distance on the data set. The set of immediate neighbors of a position n_i on the grid is denoted by $N(i)$.

The clustering of ESOM is based on the U*C clustering algorithm described by [23]. Consider a data point x at the surface of a cluster C, with a best match of $n_i = bm(x)$. The weight vectors of its neighbors $N(i)$ are either within the cluster, in a different cluster or interpolate between clusters. Assume that the inter cluster distances are locally larger than the local within-cluster distances, then the U-heights in $N(i)$ will be large in such directions which point away from the cluster C. Thus, a so-called immersive movement will perform to lead away from cluster borders. This immersive movement is performed which starts from a grid position, keeps decreasing the U-matrix value by moving to the neighbor with the smallest value, then keeps increasing the P-matrix value by moving to the neighbor with the largest value. The details of this clustering algorithm can be referred to [23].

III. METHODOLOGY AND EXPERIMENT

We have collected a total of 46 palsy images and 21 normal images. The palsy images were obtained from Mater Misericordiae Health Services in Brisbane. These images show facial weaknesses, such as not being able to close both eyes, twitching of the corner of mouth when smiling, and asymmetry cheeks. The normal images were obtained from Asian Emotion Database [24]. The database is arranged into train and test sets separately. We first grade the images manually into 5 grades. Due to the insufficient number of images, we are not able to obtain grade VI images.

Figure 1 illustrates the overall process of the automated House-Brackmann Grading system for face images. To remove noises, Gaussian blur is applied on the sample face images. The images are then converted into gray scale images and normalized using a Histogram equalization process.

An automated face-area detector is developed using the Haar Cascade classifier of OpenCV (Open Source Computer Vision: http://opencv.org) API. The face-area detector automatically locates a face in a sample image and places a region of interest (ROI) rectangle. The detector is calibrated such that it detects an ROI rectangle area that encloses the far edges of the eyes, the top of the eye brows, and the bottom of the lips. For the training data sets, the rectangle is manually rotated and/or moved to center the faces inside the rectangle such that the center line of the rectangle passes through the center of the face, i.e., between the eyes and the center of nose.

Table 2. Performance Metrics Using Emergent Self-Organizing Map

Grade	Generalization	Recall	F-Measure	GMts	GMtr
I.	59.38%	91.43%	72.00%	59.54%	93.54%
II	93.75%	97.14%	95.42%	80.23%	98.37%
III.	87.5%	100.00%	93.33%	62.06%	100.00%
IV.	90.63%	100.00%	95.08%	56.73%	100.00%
V.	68.75%	97.14%	80.52%	69.63%	95.74%

The rectangle regions of the face images are then extracted and down sampled to 100×100 pixel gray-scale images.

Figure 2 shows the intermediate images of normal face. Figure 3 shows intermediate palsy images with asymmetry mouth. For ethical reasons, these images cannot be published. Thus, sample images from the Internet of facial palsy patients are used instead. The face edges are extracted using Sobel edge detector. Figure 2(c) and Figure 3(c) show the differences between left and right faces when we fold the faces symmetrically. The differences are measured using Hamming Distance [25], i.e., the number of positions at which the corresponding pixels are different. Therefore, each image is represented by 100 Hamming distance values.

Principal component analysis (PCA) is also performed on the Hamming distance features and the top-ten principal components are selected as another set of features of the sample data sets. To compute the principal components, the Hamming distance features of face images are represented in a 100 by 67 matrix (100 rows and 67 columns), where each column is the Hamming distance features of a face image. Eigenvectors are computed using OpenCV API. The top ten eigenvectors are selected by sorting them in descending order of their eigenvalues. The 5 groups of images are then classified using 5-ary classifiers.EXPERIMENTAL RESULTS

This section showcases results of experiments. In a Receiver Operating Characteristic (ROC) curve, the sensitivity (true positive rate) is plotted as a function of specificity (false positive rate) for different cut-off points. A test with perfect discrimination (no overlap in the two distributions) has a ROC plot that passes through the upper left corner. Therefore the closer the ROC plot is to the upper left corner, the higher the overall accuracy of the test [26]. Figure 4(a) shows the ROC curve of SVM with ten principal components, the area under the curve is 0.73. Figure 4(b) shows the ROC curve of Hamming distance with SVM which gives better results of about 0.87. It provides good curve for palsy data as shown.

Accuracy may not be a sufficient measure to evaluate the performance of classifiers. A classifier may be able to obtain a very high accuracy by classifying all instances to majority class but classifying wrongly the minority class. Therefore, the use of accuracy and mean square error rate are inappropriate for imbalanced dataset like this emotion dataset. So we adopted the use of other kinds of evaluation metrics to measure the classifiers' capability to differentiate the two-class problem under the case of imbalanced data distribution. They are namely, F-measure and Geometric Means Measure (GMM), in which the performance measures are independent to prior probabilities. F-measure is a metric derived from recall and precision. Some variants would make the weighting equal as it is considered that the occurrences of false positive and false negative are likely equal. According to the evaluations by Barandela et al. [27], GMM is a more appropriate metric to evaluate the classifier performance on the datasets which are imbalanced distribution. Both F-measure and GMM are good indicators as they maximize the accuracy on each of the two classes while keeping these accuracies balanced. The GMM is defined as the square root of the product of accuracy on positive samples and negative samples. Table 1 shows the results of using ESOM classifier on the principal component features. The results are encouraging; the F-measure is above 72%. Especially in the case of Grade II, III, and IV, the results are above 90%.

V. Conclusions

In this paper, we propose to examine facial nerve palsy using SVM and Emergent Self-Organizing Map (ESOM) and grade the degree of nerve damage using House-Brackmann score. Traditional evaluation method involves medical doctor taking a thorough history of patient and determines the onset of paralysis, the rate of progression and etc. The use of automated grading system greatly reduces assessment time and increases consistency as references of all palsy images are stored to provide comparison. The results are very encouraging. We showed that by using face symmetry Hamming distance combined with Emergent Self-Organizing Map, we are able to achieve classification rate of up to 95%.

Our future works includes implementing this work to cell-phone as a mobile facial palsy assessment tool. This can provide timely diagnosis to patients situated far from medical centers. Unlike existing medical diagnostics, such as contact thermal imaging, our method can utilize existing cell phones and infrastructure to provide a robust, noninvasive, and rapid assessment method based on the simple analysis of images. The developed system can be rapidly deployed to many needed communities over the Internet. In five years, most Africans will have cell phones. Our low-cost cell phone-based data collection and diagnostics systems can be used to construct a world-wide database of diagnostic images and cases to discover the causes of diseases and patterns of disease development in developing countries.

VI. Acknologyment

This work was supported by JCU Research Grant JCUS/003/2011/IS, a grant from the Bill & Melinda Gates Foundation through the Grand Challenges Explorations Initiative (Grant Number: OPP1032125), and T-BOP PTE LTD.

References

[1] C. A. Kasse, R. G. Ferri, E. Y. C. Vietler, F. D. Leonhardt, J. R. G. Testa, and O. L. M. Cruz, "Clinical data and prognosis in 1521 cases of Bell's palsy," *International Congress Series,* vol. 1240, pp. 641-647, 2003.

[2] MedicineNet.Inc. (1996). *MedicineNet.com*. Available: http://www.medterms.com/script/main/art.asp?articlekey=6556

[3] S. Ghali, A. MacQuillan, and A. O. Grobbelaar, "Reanimation of the middle and lower face in facial paralysis: Review of the literature and personal approach," *Journal of Plastic, Reconstructive & Aesthetic Surgery,* vol. In Press, Corrected Proof, 2010.

[4] A.D.A.M. Medical Encyclopedia. (2010). *Bell's palsy*. Available: http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001777/

[5] J. P. Browder, "Facial paralysis in children," *Journal of Ear Nose Throat* vol. 57, pp. 278-83, 1978.

[6] J. W. House and D. E. Brackmann, "Facial nerve grading system," *Otolaryngol Head Neck Surg,* vol. 93 pp. 146-147, 1985.

[7] J. T. Vrabec, D. D. Backous, H. R. Djalilian, P. W. Gidley, J. P. Leonetti, S. J. Marzo*, et al.*, "Facial Nerve Grading System 2.0," *Otolaryngology - Head and Neck Surgery,* vol. 140, pp. 445-450, 2009.

[8] C. Chang and C. Lin, *{LIBSVM}: a library for support vector machines*, 2001.

[9] A. Ultsch and M. F., "ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM," University of Marburg Dept. of Mathematics and Computer Science, Germany2005.

[10] B. Lei, I. Song, and S. A. Rahman, "Optimal watermarking scheme for breath sound," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1-6.

[11] I. Song and N. V. Marsh, "Anonymous Indexing of Health Conditions for a Similarity Measure," *Information Technology in Biomedicine, IEEE Transactions on,* vol. 16, pp. 737-744, 2012.

[12] J. Vong, J. Fang, and I. Song, "Delivering financial services through mobile phone technology: a pilot study on impact of mobile money service on micro–entrepreneurs in rural Cambodia," *International Journal of Information Systems and Change Management,* vol. 6, pp. 177-186, 2012.

[13] V. Vapnik and A. Lerner, "Pattern Recognition using Generalized Portrait Method," *Automation and Remote Control,* vol. 24, 1963.

[14] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. Fifth Annual Workshop on Computational Learning Theory*, 1992, pp. 144-152.

[15] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning,* vol. 20, pp. 273-297, 1995.

[16] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery,* vol. 2, pp. 121-167, 1998.

[17] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," *Machine learning: ECML-98,* pp. 137-142, 1998.

[18] J. D. M. Rennie, "Improving multi-class text classification with naive Bayes," Massachusetts Institute of Technology, 2001.
[19] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE,* vol. 78, pp. 1464-1480, 1990.
[20] A. Ultsch, "Self-organizing neural networks for visualization and classification," 1993.
[21] A. Ultsch, "Maps for the visualization of high-dimensional data spaces," in *Proc. Workshop on Self organizing Maps*, 2003, pp. 225-230.
[22] A. Ultsch, *U*-matrix: a tool to visualize clusters in high dimensional data*: Fachbereich Mathematik und Informatik, 2003.
[23] A. Ultsch, "Clustering with SOM: U*C," presented at the WSOM, 2005.
[24] J.-J. Wong and S.-Y. Cho, "A Support Vector Reduced Multivariate Polynomial Model for Face Emotion Recognition," *IEEE Transactions on Information Forensics and Security,* 2007.
[25] Hamming and R. W., "Error detecting and error correcting codes," *Bell System Technical Journal,* vol. 29, pp. 147-160, 1950.
[26] M. Maloof, "Learning when data sets are imbalanced and when costs are unequal and unknown," in *Proceedings of the ICML, Workshop on Learning from Imbalanced Data Sets*, 2003.
[27] R. Barandela, J. S. Sánchez, V. García, and E. Rangel, "Strategies for learning in class imbalance problems," *Pattern Recognition,* vol. 36, pp. 849-851, 2003.