IoT BASED REAL-TIME VOICE ANALYSIS AND SMART MONITORING SYSTEM FOR DISABLED PEOPLE

Ghazanfar Latif [a], Adil H. Khan [b], M. Mohsin Butt [c], Omair Butt [d]
[abd] Prince Mohammad bin Fahd University, Khobar, Saudi Arabia
[c] King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
Corresponding email: glatif@pmu.edu.sa

Abstract

This research emphasizes on Internet of Things (IoT) based affordable platforms to take proper and timely measures for disabled people. It is usually observed that people with different disabilities face difficulties in all walks of life, and adequate caretaking measures are not adopted in most cases. Real time and consistent caretaking for such handicapped people is a tedious task. This paper introduces an IoT based real time analysis and alerting system for the disabled people. The proposed standalone system consistently monitors voice activity of person and in case of any abnormality in analysis outcomes, the system automatically notifies concerned hospital or caregiver to prompt for the patient's situation. The voice features are extracted from analysed voice by employing Discrete Cosine Transform (DCT), and classified through Support Vector Machine (SVM). The prototype has been developed by using Raspberry Pi single board along with voice recording module, Wi-Fi module and LCD Screen. Cloud web services have been used to store the real time activity and performing voice analysis. Montreal Affective Voices (MAV) dataset has been utilized for training and testing of voice recognition. The designed system can be regarded as a rescue system for people suffering from various life threatening health conditions including bipolar disorder, hysteria, cardiac arrest, etc. An accuracy of 81.74% has been achieved for MAV dataset, whereas an accuracy of 67.90% is achieved for real time voice input as depicted in the analysed results.

Keywords: Internet of Things (IoT), Voice Analysis, Monitoring System for Disables, DCT Features, Support Vector Machine (SVM).

1. Introduction and Background

A considerable percentage of humans suffers from diverse types of health disorders including mental and physical deficiencies. One of the commonly observed challenges is to ensure prompt and effective measures for certain fatalities. In order for the caregiver the timely reaction is only possible if a timely and accurate report is received. Hence, a continuous monitoring procedure is required after identification, which is likely to be possible only at hospital or healthcare centre. People need to wait for long time for hospitalization due to less number of beds and hospitals for people with disabilities (Torrey, et al., 2015).

With the advent of new technologies including Internet of Things, Bigdata and cloud computing, developing fully automated embedded systems are becoming a trivial task (Gubbi, et al., 2013). Internet of things (IoT) is affecting and changing the way we work, live, drive, communicate,

Page 191

ISBN:978-0-995398-016
www.apiar.org.au

purchase, sell, remember, etc. It is and will have a huge impact on evolving many sector to unprecedented levels. Marketing will be individualized according to each person, his/her behaviour and his/her purchasing habits. Cars and transportation are moving rapidly towards automatic driverless and supposedly accident free transportation. Education is evolving to virtual learn anytime anywhere concept. The new computer technologies, healthcare information management systems and more powerful small sized low price computing devices, computer science and information technology also get involved to improve healthcare sector (Feng, et al., 2013). Now days, technology is playing any important role to improve our lives in every domain.

The important factor to be considered is the cost incurred for caregiving facilities either by hospitalization or by a home care nurse (Bardhan, et al., 2013). An IoT based system seems to be a convenient and economic solution for such issues. It can aid in timely detection of a health issue without any human intervention and also reduce the healthcare cost tremendously (Gaynor, et al., 2014). Moreover, the over crowdedness at hospitals can be minimized by need based treatments at hospitals or to accommodate healthcare facility at home in nonlife threatening conditions (Kao, et al., 2015). Such systems would also aid in reducing the stress of relatives, and keeping them aware in case any assistance is required. The acquired data stored in the cloud would be a means for doctors to examine patient's health (Chauhan, et al., 2014).

3. Proposed System

The proposed system composed of four parts, including voice analysis module, features extractor/ emotional disorder classifier, notification module and a cloud data centre. In the initial phase, the voice signals are captured by a small sized wearable device and based on the frequency value the system decides to either pass the samples for classification or to discard it. If the set requirements are met, the analysed voices samples are forwarded to feature extractor and further classification is done for analysing the emotional disorder. In the third phase, alert notifications are sent to the caregiver in case of diagnosed disorder and finally the captured information is transferred to the cloud server which can be used for long term and short term analysis and decision making by the doctors. Figure 1 show the workflow process of the proposed system.

3.1 IoT based Wearable Embedded System Prototype

The proposed system prototype has been implemented using Raspberry Pi, having an LCD Touch Screen and Voice Recording Module. Raspberry Pi is an economic and low weight powerful single board processing board, specifically designed for IoT based products design and testing. It has a 1 GB memory, 1.2 GHz Quad Core processing power (Upton et al., 2014). Python script has been used to enable the voice recording module and integration of the LCD touchscreen with Raspberry Pi (Smith, et al., 2013). Power bank is used as power source. The system architecture is presented in Figure 2. A Linux based operating system Raspbian can be copied in memory card and it start up just as plug and play device. The Raspbian operating system contains more than 35,000 packages and pre-compiled software's.

The system is capable to operate in two modes. Primarily the unmanned mode is active, in which the Voice Recorder Module remains active consistently. It creates sample chunks of 15 seconds for analysis of voice frequency and compares them to a threshold for decision making. The decision module either enables feature extractor or discards the analysed data. As a secondary option the patient can dial an emergency contact using the LCD Touch Screen available in the system.

Page 192

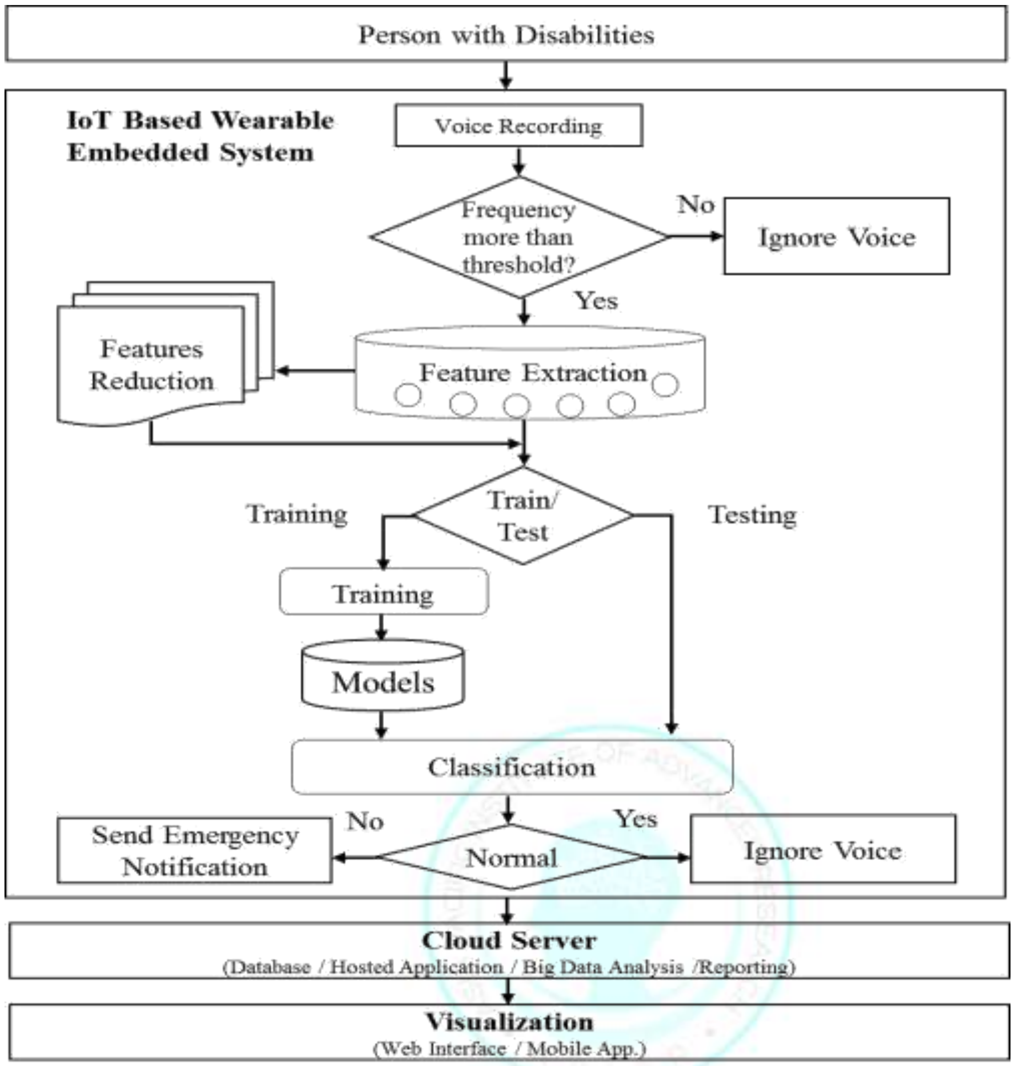

Figure 1. Workflow of the Proposed System

Figure 2. Components of IoT based wearable Embedded System. From left to right: A. Wrest case, Touch Screen, Raspberry Pi, Voice Recorder Module

Page 193

As per the defined procedure, upon classification of features showing abnormality, an emergency message will be sent to the caregiver. The classified voice details are also transferred to the cloud server for a secondary observation. Figure 3 presents the initial porotype testing on wrist as a wearable device.

Figure 3. Portotype of the Wearable Embedded System

3.2 Voice Feature Extraction (DCT)

In the voices analysis Discrete cosine transform (DCT) is a powerful tool used for signal and image processing which is used to transform them from special to frequency domain. DCT is used for converting the voice waveform signals into frequency components. DCT also has the property of separability and symmetry. 2-Dimensional DCT of the input is defined by (Khayam, et al., 2003; Strang, et al., 1999).

$$(,) = ()()\sum_{=0}^{-1}\sum_{=0}^{-1}(,)\,[\frac{(2\ +1)}{2}]\,[\frac{(2\ +1)}{2}]$$

Where 0≤ u ≤ N, & 0≤ v ≤ N, and

$$() = \begin{cases} \sqrt{\frac{}{}} & = 0 \\ \sqrt{\frac{}{}} & \neq 0 \end{cases}$$

In the proposed system, voice waveform features are extracted by using the DCT technique. In order to choose best size of features set to get more accuracy and by avoiding excess usage of computations power, several tests are performed with different number of features set like testing is done with top 15, 25, 35 and 50 frequency components. Figure 4 shows the original voice waveform and its DCT feature frequencies.

Page 194

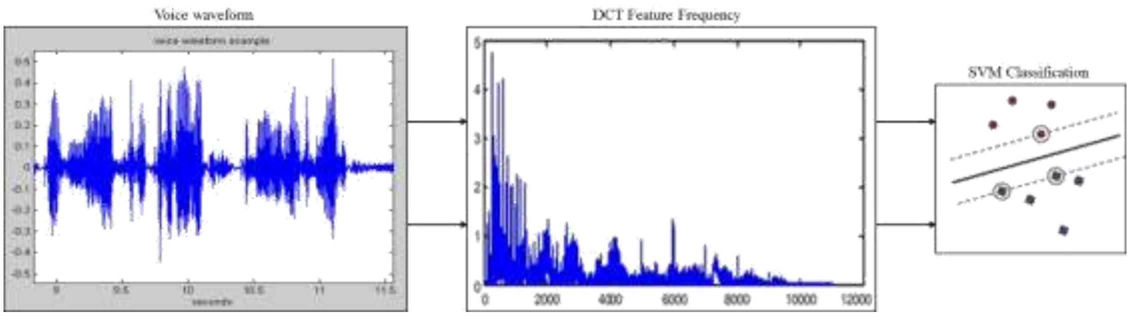

Figure 4. original voice waveform and its DCT feature frequencies

3.3 Voice Classification (Multi-Class SVM)

Support vector machine (SVM) is one of the most popular technique for classification and used to solve numerous real world problems (Brereton et al, 2010), text categorization (Furey, et al., 2000), voice recognition (Jaochims, et al., 1998) and microarray gene expression data analysis. Proposed system used SVM for multiclass classification of voice waveform.. Fig.4 shows the simple linear support vector machine. The basic theme of SVM is to maximize the margins between two classes of the hyperplane (Kumar, et al., 2015; Schölkopf, 2002).

Figure 5 shows the hyperplane and margin. Let a set of n training data of separable classes {(1, 1), (2, 2),..., (,)} ,i=1,2,...,n. where ∈ is an n dimensional space and = ±1.Given a weight vector w and bias weight b, the separation of hyperplane between multiple classes can be defined by the equation below equations. The separation of classes by equations is a linear separation. Any hyperplane can be defined by the equation.

$$(w.x_i+b) \geq 1, \quad \text{if} \quad y_i = 1$$

$$(w.x_i+b) \leq -1, \quad \text{if} \quad y_i = -1$$

$$w.x_i+b=0$$

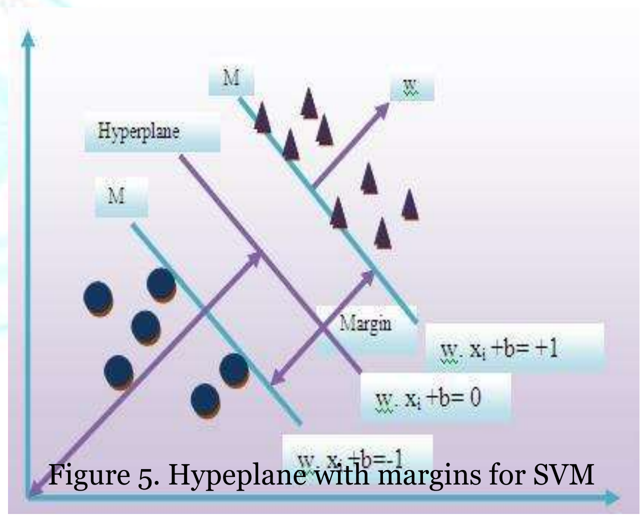

Figure 5. Hypeplane with margins for SVM

SVM tries to maximize the margin between these classes by minimizing ½||w||2. Quadratic optimization algorithms can identify which training pointsare support vectors with non-zero Lagrangian multipliers . This optimization problem can be defined by the following equation.

$$=\Sigma \quad - \quad \frac{1}{2}\Sigma\Sigma$$

These supporting vector are used for determining the decision functions, and all other data points are discarded.

Page 185

3.4 Emergency Notifications

The developed device detects any abnormality based on the voice analysis and the system will send notification automatically to the hospital personnel and their relatives to promptly send caregiver to attend the patient. Further, the system stores the patient's data in the cloud which can also be to analysed to generate reports for reviewing the person's health to make appropriate decisions.

4. Experimental Data and Results

For initial training and classification of different emotional voices, Montreal Affective Voices (MAV) dataset is used (Belin et al., 2008). It consists of total ninety nonverbal different voices corresponding to emotions of anger, disgust, fear, pain, sadness, surprise, happiness and pleasure and natural expression. The voices are being recorded by 10 different actors including five males and 5 females. Testing on realtime recorded voices also done. 27 emotion voices of 3 males are used for realtime voices classification.

Features are extracted by using DCT and multiclass SVM classifier is used with different feature set including 15, 25, 35 and 50 top features set. We compared the classification results with K-nearest neighbour (KNN) classifier. The experimental results in Table 1 proves that multiclass SVM gives best accuracy results by using 25 DCT features set.

Table 1. Comparison of both Classification Techniques
with different number of features set of MAV dataset

Number of Top DCT Features Used	**Multiclass SVM**		**KNN**	
	Accuracy FP Rate	Mean Absolute Error	Accuracy FP Rate	Mean Absolute Error
15 Features	76.2 %	11.80 %	63.5 %	14.83 %
25 Features	81.7 %	9.23 %	73.4 %	12.54 %
35 Features	80.4 %	10.58 %	68.3 %	13.30 %
50 Features	69.1 %	12.77 %	59.4 %	22.53%

Table 2 describes the detailed accuracy of voice classification for MAV dataset and real-time dataset by using true positive rate, false positive rate, precision and recall (Davis et al., 2006).

Table 2. Detailed Accuracy for Voice Classification by using MAV Dataset and Real-time Dataset

	MAV Voice Dataset				**Real-time Voice**			
	TP Rate	FP Rate	Precision	Recall	TP Rate	FP Rate	Precision	Recall
Anger	0.741	0.021	0.889	0.741	0.556	0.034	0.714	0.556
Disgust	0.896	0.020	0.929	0.896	0.833	0.034	0.789	0.833
Fear	0.63	0.029	0.735	0.63	0.714	0.049	0.625	0.714
Happiness	0.861	0.068	0.705	0.861	0.714	0.082	0.5	0.714
Neutral	0.875	0.037	0.892	0.875	0.571	0.049	0.571	0.571
Pain	0.791	0.066	0.72	0.791	0.5	0.048	0.5	0.5
Pleasure	0.813	0.034	0.866	0.813	0.667	0.025	0.769	0.667
Sadness	0.75	0.051	0.812	0.75	0.556	0.042	0.667	0.556
Surprise	1	0.031	0.86	1	1	0.03	0.5	1
Average	0.817	0.0396	0.823	0.817	0.679	0.0436	0.625	0.679

Conclusions

In this research, an innovative method is proposed for monitoring of disabled people based on real-time voice analysis and automatically notifying the caregivers in case of emergency. The experimental results of voice classification have shown high level of accuracy. Human motion monitor is a medical device whose intended use is to diagnose the disorders, or in the cure, treatment, or prevention of diseases. The emotional voice data can also be synchronized with the cloud for short term and long term monitoring which can further be used for decision making. Hence, the doctor and caregiver can monitor their patients from the cloud hosted system any time and any location. If the system records sudden abnormal activity or behavior by the patient, it will automatically send an alert message to the emergency or to the patient's caregivers.

References

i. Bardhan, I. & Thouin, M., 2013. Health information technology and its impact on the quality and cost of healthcare delivery. *Decision Support Systems,* 55(2), pp. 438-449.

ii. Belin, P., Fillion-Bilodeau, S. & Gosselin, F., 2008. The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods,* 40(2), pp. 531-539.

iii. Brereton, R. & Lloyd, G., 2010. Support vector machines for classification and regression. *Analyst,* 135(2), pp. 230-267.

iv. Chauhan, R. & Kumar, A., 2013. *November. Cloud computing for improved healthcare: techniques, potential and challenges. In E-Health and Bioengineering Conference (EHB).* s.l., IEEE.

v. Davis, J. & Goadrich, M., 2006. *The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine learning (pp. 233-240).* s.l., ACM.

vi. Feng, D., Siu, W. & Zhang, H., 2013. *Multimedia information retrieval and management: Technological fundamentals and applications*. 1st ed. s.l.:Springer Science & Business Media.

vii. Gaynor, M., Wyner, G. & Gupta, A., 2014. *Dr. Watson? Balancing automation and human expertise in healthcare delivery. In International Symposium On Leveraging Applications of Formal Methods, Verification and Validation*. Heidelberg, Springer Berlin.

viii. Gubbi, J., Buyya, R., Marusic, S. & Palaniswami, M., 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems,* 29(7), pp. 1645-1660.

ix. Jaochims, T., 1998. *Text categorization with support vector machines. In Proceedings of the European Conference on Machine Learning*. s.l., Springer-Verlag.

x. Kao, C., Yang, J. & Lin, C., 2015. The Impact of Ambulance and Patient Diversion on Crowdedness of Multiple Emergency Departments in a Region. *PloS One,* 10(12), pp. 144-227.

xi. Khayam, S., 2003. *The discrete cosine transform (DCT): theory and application.* Michigan: Michigan State University.

xii. Kumar, S. & RangaBabu, T., 2015. Emotion and Gender Recognition of Speech Signals Using SVM. *Emotion,* 4(3).

xiii. Schölkopf, B. & Smola, A., 2002. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. s.l.:MIT Press.

xiv. Smith, B., 2013. *Raspberry Pi Assembly Language RASPBIAN Beginners: Hands On Guide*. s.l.:CreateSpace Independent Publishing Platform.

xv. Strang, G., 1999. The discrete cosine transform. *SIAM Review,* 41(1), pp. 135-147.

xvi. T.S.Furey, et al., 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics,* 16(10), pp. 906-914.

ISBN:978-0-995398-016
www.apiar.org.au

xvii. Torrey, E. et al., 2015. The shortage of public hospital beds for mentally ill persons. *Montana,* 303(29), pp. 6-9.

xviii. Upton, Eben & Halfacree, G., 2014. *Raspberry Pi user guide.* s.l.:John Wiley & Sons.