Smartphone-Based Gait Assessment to Infer Parkinson's Disease Severity using Crowdsourced Data

Hamza Abujrida, Emmanuel Agu, *Member, IEEE*, Kaveh Pahlavan, *Fellow, IEEE*

Abstract— **People afflicted with Parkinson's Disease (PD) experience impairment of their gait (the way a person walks), which frequently results in falls. In this paper we investigate a machine learning method to assess PD severity using accelerometer data passively crowdsourced from participants' smartphones while they walked. Time and frequency domain features such as entropy rate and peak frequency, and postural sway features were extracted from accelerometer data and classified. Our work is the first to classify PD severity on the UPDRS scale and distinguish PD patients from controls, using noisy crowdsourced data. Our crowdsourcing approach examined 50 patients in the wild, demonstrating the potential to use smartphone sensing to remotely assess and monitor PD patients at the population level. The random forest classifier was the most accurate in distinguishing subjects from controls with an average accuracy of 87.03% and also for assessing PD severity (Normal, Slight, Mild, Moderate and Severe), with an average accuracy of 85.8%.**

I. INTRODUCTION

Parkinson's disease (PD) is a neurological disorder, which reduces a patient's ability to regulate their movements, emotions, and cognitive functions (memory and concentration loss). Approximately 630,000 people in the United States were diagnosed with PD in 2010, which could double by 2040. PD causes shakiness in patients, tremors and stiffness when moving their limbs, bradykinesia (slowness of movement), FoG (Freeze of Gait), Micrographia (Small hand writing) and problems with balance and falls as the disease progresses. Complications due to PD are the 14th leading cause of death in the US [2]. PD dramatically alters the patient's gait, which we focus on in this work.

PD treatment costs exceeded $14.4 billion in 2010 (about $22,800 per patient) [12]. Majority of PD-associated costs arise from hospital visits for assessment and in-patient days. Assessments include patients' walking and balance using standardized measures such as the Movement Disorders Society Unified PD Rating Scale (MDS-UPDRS), which is the most widely used objective clinical rating scale for PD. Remote monitoring using electronic devices such as smartphones enable caregivers to adjust a patient's medication dynamically, and minimize unnecessary hospital visits. Additionally, compared with in-clinic monitoring, remote monitoring enables capturing longitudinal data in realistic contexts for longer periods. For instance, patients' physiological response after taking their medication can be actively monitored. As smartphones are now owned by over 77% of Americans, PD monitoring using smartphones can potentially have a large impact.

The MDS-UPDRS [15] has four parts: 1) Non-motor experiences of daily living 2) Motor experiences of daily living 3) Motor examination, and 4) Motor complications. In this paper we focus on part 3 of the MDS-UPDRS [15], specifically using a machine learning approach to infer PD severity and also to discriminate between the gait (walk) of PD subjects and controls by classifying features extracted from accelerometer data passively gathered from their smartphones. We utilize statistical, time, frequency and statistical domain features as well as postural sway features, which are widely used in clinical settings for gait analysis.

The majority of prior smartphone-based PD studies have been conducted with relatively few (10-20 participants), which increases the chance of classification errors due to overfitting when using a machine learning approach. Additionally, most studies were conducted in highly controlled settings, typically under the guidance of a human proctor. In order to gather lots of data quickly, at low cost with minimal constraints, crowdsourcing has been proposed for conducting evaluations and scientific experiments [16]. In healthcare, crowdsourcing has been explored for studies in healthcare provider decision making [17], voice-based diagnostics for depression [18] and skin self-examination [19]. For chronic diseases such as PD where large scale data collection from Patients With Parkinson's (PWP) is extremely challenging, crowd-sourcing can facilitate quick and inexpensive data collection. However, crowdsourced data can be noisy with recording errors and missing or inconsistent entries. In fact, to the best of our knowledge, it is unknown whether accurate machine-learning based remote assessment of PD patients can be based on crowd-sourced data.

In this paper, we explore remote assessment of PD patients using crowdsourced data from the mPower study [4], a clinical observational PD study conducted purely using a Smartphone app (Figure 3). The study repeatedly interrogated aspects of the PD movement disorder through EMA-style surveys and continuous accelerometer recordings from 9520 participants with and without Parkinson disease [4] between March-September 2015. Using the mPower dataset, we explored the utility of crowdsourced patient self-assessments with minimal participation by clinicians or caregivers. Our methodology which gathers accelerometer data passively from a smartphone placed in the patient's front pocket while they are walking imposes a minimal burden on patient and paves the way for remote PD assessment and patient follow-ups.

Prior analytical approaches to assessing PD tremors include analysis in the time domain [7], frequency domain [8], and a combination of time, frequency and nonlinear domains [9]. FoG was analyzed using accelerometer and gyroscope data from different body locations using the AdaBoost.M1 machine learning classifier [10]. Real-time FoG events were also detected from 3 external accelerometer sensors and a Smartphone [11], and Random Forest and AdaBoost yielded the best results. Majority of these studies are based on small numbers of participants, which makes their results statistically insignificant, possibly overfitted, were frequently done in controlled experiments, and also did not analyze the walking balance of participants. We address all these issues in this paper by utilizing crowdsourced data from 50 participants. The random forest classifier was the most accurate in distinguishing subjects from controls with an average accuracy of 87.03% and also for assessing PD severity (Normal, Slight, Mild, Moderate and Severe), with an average accuracy of 85.8%. Random Forest performed well due to its ability to identify non-linear relationships in the feature space.

978-1-5386-1392-4/17/$31.00 ©2017 IEEE

II. BACKROUND: THE MPOWER STUDY

The mPower Study [4] enrolled 9520 individuals who were 18 years or older and had been diagnosed with PD or participated as controls (healthy subjects). Participants initially completed a baseline/demographics survey indicating whether they had been professionally diagnosed with PD. Surveys were completed monthly including a subset of the MDS-UPDRS shown in Table 1, including questions about the motor symptoms of PD, a walking balance value ranging from 0 (no motor impairment) to 4 (maximal impairment). Survey responses represent self-reported outcomes and thus occasionally contain typographic errors and possibly inconsistent information. Participants were asked to perform 4 types of activities (memory, tapping, voice and walking). Participant were asked to perform the walking activity 3 times a day: 1) Immediately before taking their medication, 2) after taking their medication 3) at some other time. Our work focused on the walking activity performed before taking the medication to avoid its effects [6]. Participants were instructed to walk 20 steps in a straight line, stand still for 30 seconds and then walk back. Their walk accelerometer data was recorded by a smartphone app and later used to evaluate their gait and balance. We analyze the first 20 steps on the outbound walk to avoid complications such as Freezing of Gait (FoG), which usually occurs when patients stand still or turn around while walking.

III. METHODOLOGY

Accelerometer Posturography Features (APF) [5] have long been used for measuring subtle balance deficits in clinical settings [20]. In our work we seek to evaluate the effectiveness of APF in discriminating between a normal vs PD gait pattern and classify the walking balance of PD participants. Accelerometer data collected from 50 participants (28 Subjects and 22 controls) as they walked 20 steps (20-30 seconds) in a straight line while a Smartphone was placed in their front pocket. APF which are sensitive to PD gait problems and other gait features were extracted and classified using a machine learning approach illustrated in figure 1. The accuracy of different Classification algorithms was compared (see results).

A. The mPower Dataset and Patient Filtering Rules:

In the mPower Walking dataset, 6805 of the 9520 participants completed the enrollment demographics survey, 1087 subjects self-identified as having been professionally diagnosed with PD while 5581 were healthy controls. However only 898 participants contributed data for five days or more [4]. Participants used mostly iPhones. As the severity of walking imbalance can vary with time, as well as the patient's age, their stage of PD, and medication taken, only walking activity that took place within 3-4 weeks of reporting the severity of walking balance (filling the UPDRS survey) was considered. Subjects who recorded less than 20 steps or had inconsistent periods of walking were also eliminated. We filtered 50 participants (28 Subjects and 22 controls) who had an adequate number of walking activities (10 to 60) yielding 50 Participants (32 Male, 18 Female, mean age of 57.6 (std 15.9).

B. Feature Extraction

The smartphone accelerometer data was divided into equal 5-second segments. Features were extracted from each segment, including statistical features such as standard deviation, average step time and average cadence. Frequency domain features extracted included windowed energy, peak frequency and spectral centroid, wavelet entropy, radio spectral peak. Postural sway features such as sway area around the 3 body axes were also extracted. These features measured fluctuations in the participant's gait and body motion especially their balance while walking. Table III lists all the features extracted, the accelerometer attribute they measure, and cites examples of other work where those features were used for gait analysis. Essentially, these features quantify gait frequency, amplitude and rhythmicity, which are the movement attributes visually assessed by neurologists during the UPDRS Part III Test.

Table 1 Parts of the MDS-UPDRS in the Survey

Question	**Variable Name**	**Variable Detail**
Over the past week, have you usually had shaking or tremor?	MDS-UPDRS2.10	one of: {'Normal', 'Slight', 'Mild', 'Moderate', 'Severe'} mapping to {0, 1, 2, 3, 4}
Over the past week, have you usually had problems with balance and walking?	MDS-UPDRS2.12	one of: {'Normal', 'Slight', 'Mild', 'Moderate', 'Severe'} mapping to {0, 1, 2, 3, 4}
Over the past week, on your usual day when walking, do you suddenly stop or freeze as if your feet are stuck to the floor?	MDS-UPDRS2.13	one of: {'Normal', 'Slight', 'Mild', 'Moderate', 'Severe'} mapping to {0, 1, 2, 3, 4}

C. Machine Learning Classification

Using machine learning classification algorithms including Random Forest, Bagged Trees, SVM and KNN, we sought to discriminate participants with PD from controls and predict their PD severity from smartphone accelerometer features. Statistical analysis and classification of the smartphone accelerometer features were done using MATLAB. Machine learning (ML) classification models were synthesized to classify patients using the UPDRS 2.10 question (balance while walking) as the independent response variable.

The performance of classification models for the UPDRS III were evaluated using 10-fold cross-validation. During each repetition of 100 folds for cross validation, the following algorithm was used to test the model:

1) Use stratified sampling to split the data into 90% training and 10% testing subsets,
2) Train the classifier using the training set.
3) Predict the class the features into two sets of bins defined as:
   a) Professionally diagnosed PD patient vs Controls (2 classes)
   b) Walking Balance Severity (5 classes of Walking Balance for PD patients (As in table 1, UPDRS III 2.10).
4) Compute classifier performance metrics such accuracy, F-measure, recall, precision and confusion matrix.

IV. RESULTS

Below we present our results. First, we generated box plots to visually examine the sensitivity all the features to PD gait imbalances. Figure 2 shows the Average Step Time (AST) feature. Clearly, AST increased as walking balance worsened

A. Feature Selection & Removal of Redundant Features

We selected 16 features from our initial list of 27 features by applying two criteria. First, we selected accelerometer features that were highly correlated with gait imbalances. Additionally, we reduced multicollinearity of features [21] by computing a correlation matrix of all features and removing features highly correlated with each other (correlation of 0.75 or higher). Correlation was calculated using Pearson's coefficient:

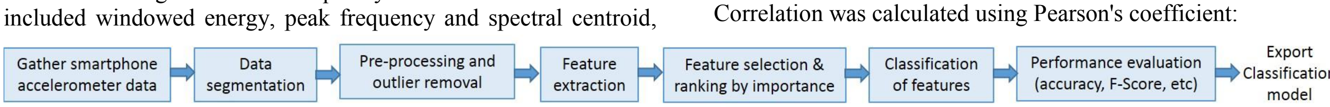

Figure 1 Flow Diagram for data collection, feature extraction and classification

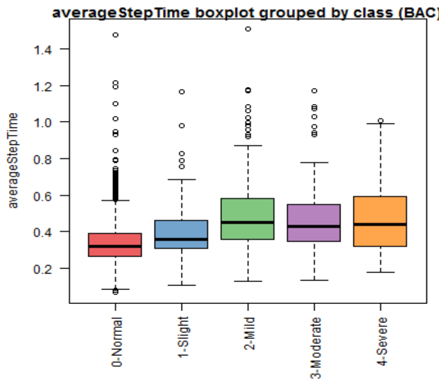

Figure 2 Box Plot showing AST value becomes less as the Severity of balance increase.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

where $\bar{x}$ and $\bar{y}$ are the sample means of X and Y, and s_x and s_y are the corrected sample standard deviations of X and Y.

The average cadence, standard deviation, RMS average power, energy in 5 to 3 Hz, spectral centroid, wavelet entropy, radio spectral peak FFT, radio spectral peak DCT, average step length, and gait velocity features were removed due to high collinearity. The features number of steps, average step time, skewness, kurtosis, min-max difference, coefficient of variation of step time, harmonic ratio, cross correlation, entropy rate, radio spectral peak, SNR, THD, windowed energy in 5 to 3Hz, peak frequency, bandwidth and wavelet band were used for training and testing the ML classification models.

B. *Ranking Features by Importance*

The importance of features indicates the relative contribution of each feature to the accuracy of the classifier. While constructing a decision tree, a variable's importance can be calculated as the decrease in the prediction error (Mean Squared Error - MSE) when the decision tree is split by that variable. The importance of our selected features is plotted in figure 4. Average Step Time, entropy rate and peak frequency were the 3 most important features, while Total Harmonic Distortion and harmonic ratio were least important.

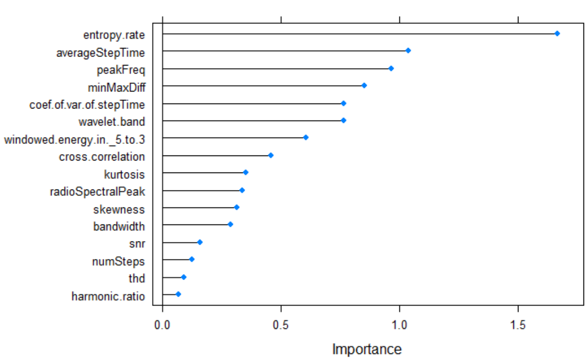

Figure 3 Ranking of Features by Importance

A. *Results of classification.*

The selected features were classified into the walking balance classes (Normal, Slight, Mild, Moderate, Severe). The Accuracy of various classification algorithms were compared including bagged trees, SVM, KNN and random forest. Two main classification experiments were explored. First, the accelerometer features were classified 1) to distinguish subjects with PD from controls (2 bins) and 2) to classify PD severity based on subjects' walking balance. Additionally, various numbers and sizes of folds were explored. Figure 4 summarizes the main results:

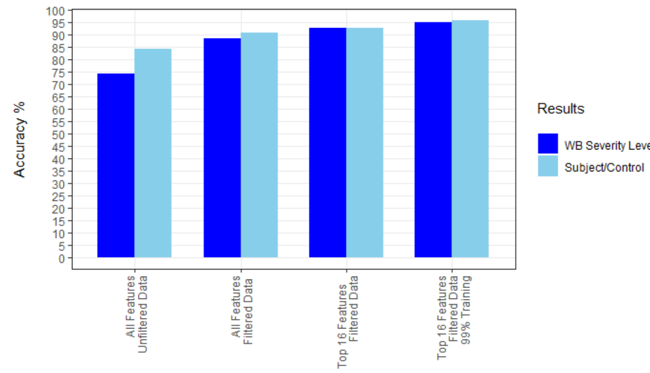

Figure 4 Bar Chart of Classification Results (For the first 3 bins Training set is 90%).

1- All features extracted from all data unfiltered: Random forest (average, highest) classification accuracy were (68%,74%) for PD severity and (79%, 84%) for PD subject vs control.

2- All features extracted from filtered data: (Removed walks less than 10 seconds and subjects whose accelerometer data was recorded more than 4 weeks after their survey response). Random forest (average, highest) accuracies were (85%,88%) for PD severity and (85%,90%) for PD subject vs control.

3- Selecting only the top 16 features by importance: Random forest (average, highest) accuracies were (86%, 92.5%) for PD severity classification and (87%, 92.5%) for PD subject vs control.

4- Using the top 16 features (99%, 1%) split: Random forest's (average, highest) accuracies were (87.5%, 94.7%) for PD severity classification and (88.9%, 95.5%23) for PD subject vs control.

5- Comparing classification algorithms: Table 2 compares the accuracies of random forest, SVM, ensemble bagged trees and KNN.

Table II Comparison of Classification Algorithms

Classifier Configuration			Results	
Attribute Classified	**Classifier**	**Test Set**	**Average Accuracy**	**Response Variable**
Accelerometer Posturography Features	Random Forest	Percentage split, 90% train 10% test	86.4%	Walking Balance
	Ensemble Bagged Trees	Percentage split, 90% train 10% test	84.1%	Walking Balance
	SVM	Percentage split, 90% train 10% test	72.4%	Walking Balance
	KNN	Percentage split, 90% train 10% test	73.2%	Walking Balance
Accelerometer Posturography Features	Random Forest	Percentage split, 90% train 10% test	87.2%	Subject/ Control
	Ensemble Bagged Trees	Percentage split, 90% train 10% test	84.1%	Subject/ Control
	SVM	Percentage split, 90% train 10% test	77.6%	Subject/ Control
	KNN	Percentage split, 90% train 10% test	76.2%	Subject/ Control

V. Conclusion

Complications caused by PD are the 14th leading cause of death in the US. In this paper, we investigate a machine learning method remotely assess PD severity based on patients' walking balance and distinguish patients with PD from healthy subjects. By classifying smartphone accelerometer features derived from data crowdsourced data. Despite challenges such as missing, inconsistent and erroneous entries, we were able to distinguish 28 PD patients from 22 controls with an 87% accuracy, and classify the PD severity with an accuracy of 85%. Random forest was the most accurate classification algorithm. Our work is the first to utilize crowdsourced gait data to distinguish people with PD and their Walking balance severity level from only smartphone accelerometer data. In future work, we plan to investigate other smartphone sensors reading such as the gyroscope, and explore how to combine walking with other activities such as voice, tapping and memory tests to predict the PD stage, in order to further improve classification accuracy.

Table III. Smartphone accelerometer features used in gait analysis

No	Feature	Description
1	**Number of Steps**	The number of steps taken in a given time interval [14]
2	**Average Step Time**	The average time elapsed for each step [13]
3	**Average Cadence**	The ratio of the total number of steps by the total time [14]
4	**Skewness**	Asymmetry of the signal distribution [13] [14]
5	**Coefficient of Variation of Step Time**	Within-subject standard deviation of the stride interval divided by the mean stride interval [13]
6	**Harmonic Ratio**	Harmonic Ratio quantifies the harmonic composition of the accelerations for a given stride via DFT [13]
7	**Average Step Length**	The average distance covered by each step
8	**Gait Velocity**	The ratio of the total distance covered by the total time [14]
9	**Minimum and Maximum Difference**	Global maximum of one step minus global minimum of one step, averaged over all steps of one subject [1]
10	**Root Mean Square**	Root Mean Square or quadratic mean is a statistical measure [1]
11	**Entropy Rate**	the uncertainty measure of the signal, and the regularity of a signal when anticipated that consecutive data points are related [1]
12	**Average Power**	the mean of the total power underneath the curve of the PSD estimate for a signal [13] [14]
13	**Ratio of Spectral Peak (with Welch, FFT and DCT)**	Ratio of the energies of low and high frequency bands [14]
14	**Signal Noise Ratio**	Power of whole signal over power of its computed noise [13]
15	**Total Harmonic Distortion**	Distortion of the whole signal compared to its harmonics [13]
16	**Energy in Band 0.5 to 3Hz**	Energy in a frequency band describes parts of distinct frequencies in the signal, and the frequency range is recommended as 0.5Hz to 3Hz [1]
17	**Windowed Energy in Band 0.5 to 3Hz**	Energy in frequency band of 5 second windows with an overlap of 2.5 seconds, windows from complete signal sequence are averaged [1]
18	**Peak Frequency**	The maximum spectral power [14]
19	**Spectral Centroid**	The frequency that divides the spectral power distribution into two equal parts [14]
20	**Bandwidth**	The difference between the uppermost and lower most frequencies/range of frequencies in the signal [14]
21	**Ratio of Spectral Peak (with FFT derivative)**	Ratio of the energies of low and high frequency bands [13] [14]
22	**Wavelet Bandwidth**	The relative energy contribution in a time-frequency band [13]
23	**Wavelet Entropy Rate**	Wavelet entropy represents signal disorder in the time-frequency domain [13]
24	**Zeroth-Lag Cross-Correlation Coefficient**	The agreement or similarity between 2 directional acceleration signals [13]
25	**Kurtosis**	The extent to which the distribution of signal amplitudes lies predominantly on the left of the mean amplitude [13] [14]
26	**Standard Deviation**	Measure for signal spreading, defined as the square of standard deviation [13] [14]
27	**Ratio of Spectral Peak (with DCT derivatives)**	Ratio of the energies of low and high frequency bands [14]

References

[1] J Klucken, *et al*, Unbiased and mobile gait analysis detects motor impairment in Parkinson's disease. PloS one, 8(2), February 2013.

[2] What is Parkinson's? Nat'l Parkinson Found. http://www. parkinson. org/understanding-parkinsons/what-is-parkinsons. Acc. Sept 9, 2017.

[3] S. Arora, V. Venkataraman, S. Donohue, K.M. Biglan E.R. Dorsey, M.A. Little, Detecting and monitoring the symptoms of PD using smartphones: a pilot study, Parkinsonism & Rel. Disorders,2015; 21(6).

[4] B. Bot, C. Suver, E. Neto, M. Kellen, A. Klein, C. Bare, M. Doerr, A. Pratap, J. Wilbanks, E. Dorsey, S. Friend and A. Trister, The mPower study, PD mobile data collected using ResearchKit, Sci. Data, 2, 2016.

[5] C. Aiello and E. Agu Investigating Postural Sway Features, Normalization and Personalization in Detecting Blood Alcohol Levels of Smartphone users, in Proc. Wireless Health 2016.

[6] A Muniz, H Liu, K. Lyons, R Pahwa, W Liu, F Nobre, J Nadal, Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic stimulation in parkinson disease on ground reaction force during gait. J. Biomech. 43, 720–726 (2010)

[7] J Timmer, C Gantert, G Deuschl, J Honerkamp, Characteristics of hand tremor time series. Biological Cybernetics **70**, 75–80 (1993)

[8] C. Riviere, S Reich, N. Thakor, Adaptive fourier modelling for quantification of tremor. J. Neuroscience Methods **74**, 77–87 (1997)

[9] H. Dai, P. Zhang and T. Lueth, Quant. Assessment of Parkinsonian Tremor Based on an Inertial Measurement Unit, Sensors J. 2015.

[10] H. Kim, H. Lee, W. Lee, S. Kwon, S. Kim, H. Jeon, H. Park, C. Shin, W. Yi, B. Jeon, and K. Park, Unconstrained Detection of Freezing of Gait in PD Patients Using Smartphone, Proc EMBC 2015

[11] S. Mazilu, M. Hardegger, Z. Zhu, D. Roggen, G. Troster, M. Plotnik, J. Hausdorff, Online Detection of FoG with Smartphones and Machine Learning Techniques, in Proc Pervasive Health 2012.

[12] S. Kowal, T. Dall, R. Chakrabarti, M. Storm, and A. Jain, The Current and Projected Economic Burden of Parkinson's Disease in the United States Movement Disorders, Issue 3, 2013

[13] E. Sejdic, K. Lowry, J. Bellanca, M. Redfern, and J. Brach, A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains. IEEE Transactions Neural Systems and Rehabilitation Engineering, 22(3), 2014. Pp. 603-612.'

[14] H Lu, J Huang, T Saha, and L. Nachman, Unobtrusive gait verification for mobile phones, In Proc. ACM ISWC 2014, pp 91-9

[15] C. G. Goetz *et al* for the Movement Disorder Society, Revision of the Unified PD Rating Scale (MDS-UPDRS): Scale Presentation and Clinometric Testing Results, Movement Disorders, 23(15), 2008.

[16] O. Alonso, Implementing crowdsourcing-based relevance experiment: an industrial perspective, Info. Retrieval, 16 (2). pp. 101–120, 2013.

[17] M. Sims, J. Bigham, H. Kautz, Marc. Halterman, Crowdsourcing Medical Expertise in Near Real Time, J. Hosp. Medicine, 9(7), 2014

[18] B. Schuller and J Pei, Using Computer Intelligence for Depression Diagnosis and Crowdsourcing, IEEE Computer, 49(7), July 2016

[19] A King, R Gehl, D Grossman and J Jensen, Skin self-examinations and visual identification of atypical nevi: Comparing individual and crowdsourcing approaches, Cancer Epidemiology, 37(6)2013.

[20] M. Mancini *et al*, ISway: a sensitive, valid and reliable measure of postural control, J. NeuroEngineering and Rehabilitation, 9:59, 2012.

[21] M Sharma and S Saha, Graph Based Approach for Minimum Multicollinearity Highly Accurate Regression Model Explaining Maximum Variability, in Proc Confluence 2014