Fatigue Detection Method Based on Smartphone Text Entry Performance Metrics

Hilal Al-Libawy, Ali Al-Ataby, Waleed Al-Nuaimy, Majid A. Al-Taee, Qussay Al-Jubouri
Department of Electrical Engineering and Electronics
University of Liverpool, Liverpool, UK
e-mail: {Abbood, aliataby, wax, altaeem, hsqaljub}@liverpool.ac.uk

Abstract— **Workplace fatigue increases the risk of injuries and other accidents, thus a growing interest in identifying early detection signs for fatigue has been recently reported in literature. Ubiquity of smartphones opens new opportunities for detecting early signs of fatigue, using daily user activities such as typing, chatting, Internet surfing and track-screen gestures. This paper presents a non-intrusive human fatigue detection method based on smartphone keyboard typing. This is achieved by using a smartphone application that records keystroke events time based on which a new application is developed to identify the user's fatigued/alert status. Text entry error rate, which is considered as a type of psychomotor measures for fatigue/alertness, is utilized as a ground-truth metric in this study. A binary classifier based on a support vector machine classifier is developed to identify the fatigue/alertness status and its performance is assessed experimentally. The obtained results have demonstrated a promising accuracy of 88.8%. This finding is expected to facilitate development of a low-cost and non-intrusive mobile instrument for fatigue/alertness detection.**

Keyword-fatigue; behavioral biometrics; smartphone; machine learning; error rate; SVM

I. INTRODUCTION

Fatigue detection is an essential requirement for safety in workspace because it can lead to accidents, injury, and loss of productivity and decreases performance. Numerous methods have been reported in literature to detect and quantify operator fatigue [1]. Thinking clearly is affected by fatigue [2], and this may result in making operators unable to assess their own level cognitive impairment. The loss of awareness of operators their required safety level is part of workplace injuries and accidents. One of many definitions of fatigue, which is related to performance impairment, is "decrements in performance on tasks requiring alertness and the manipulation and retrieval of information stored in the memory" [3].

Smartphones are quickly becoming the most popular device for personal communications. The widespread usage of these devices along with continuous advances in mobile technology and applications have made great changes to handhelds devices to become a new way of data collection and management for a wide range of applications including: fatigue detection [4], [5], workflow management [6], and mobile health applications [7] - [9]. Moreover, the computing power and storage of smartphone facilitate research challenges in real environments [10], [11].

This research was supported by a PhD fellowship from Babylon University-Ministry of Higher Education and Scientific Research, Iraq; Grant No. MOHESR-IQ-2013-1257.

Text entry is a daily task for any computer or smartphone users. In the time of using typewriters, it was easier to measure typing speed by using a stopwatch while errors were examined by hand. Nowadays, typing performance assessment becomes more and more difficult because of word-process auto-correction utility and auto-prediction tools in smartphones applications. New performance metrics (speed and accuracy) are developed depending on large experimental paired data of presented text (what participants were asked to enter) and transcribed text (typing output) [12].

A wide range of keyboards such as physical QWERTY keyboard and multi-tab keyboard (more than on character on the same key) is offered for different computer systems and mobile phones [13]. While, in current time, virtual touch screen keyboard is the most common keyboard used with smartphones. Recently, researchers have proposed text entry performance metrics to assess the efficiency of keyboard typing such as typing speed and error rate [12].

In this work, a human-fatigue detection method based on virtual keyboard timing metrics is proposed and implemented using the smartphone text entry. It utilizes several known metrics, which are previously reported in [14]; of these the holding time and digraph time are the most important. Two binary classifiers, support vector machine (SVM) and artificial neural network (ANN), are suggested to identify the fatigue/alertness status of users participated in this study.

The remainder of this paper is organized as follows: Section II overviews keystroke dynamics biometrics. Section III describes text entry accuracy metrics. Section IV describes the adopted dataset and feature extraction and labeling. The design of the system and its block diagrams are presented in Section V. Section VI presents classification results and some conclusions derived from the presented work.

II. KEYSTROKE DYNAMICS

Many metrics can be extracted from smartphone users when they interact with virtual keyboards text entry. These metrics, usually time events, are known as keystroke dynamics. An individual typing pattern can be obtained for each user. Such a pattern can hold much information relevant to the behavior of a particular user [15], [16].

The hold time (HT) and digraph time (two consecutive keys) are the basic features that are extracted from keystroke dynamics. Fig. 1 depicts an example of time features that are relevant to keystroke dynamics.

978-1-5090-5487-9/17 $31.00 © 2017 IEEE
DOI 10.1109/DeSE.2016.9

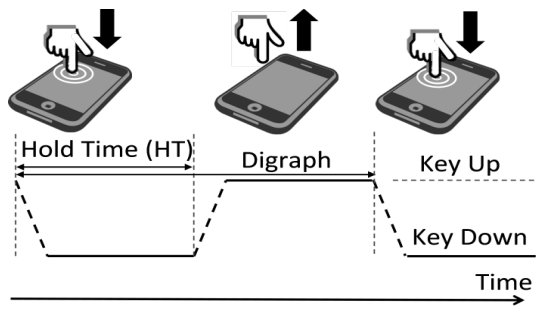

Figure 1 Graphical description of Keystroke dynamics features.

III. Text Entry Accuracy Metrics

Dealing with digital devices such as computers, tablets and smartphones is increasing dramatically. Text entry is the main interaction activity between human and these devices. Typing speed and error rate are the two major categories that mainly used to assess typing accuracy and efficiency [17].

Text entry accuracy metrics have been of interest to numerous researchers [14], [15] and there have been a common consensus on the following set of performance metrics for the typing accuracy:

- Presented Text (P) is the experiment message
- Transcribed Text (T) is the text that is written by a participant.
- Input Stream (IS) is the all keystrokes performed by the participant.
- Correct (C) is the number of the correct characters in the transcribed message.
- Incorrect Not Fixed (INF) is the number of the incorrect characters that appear in the transcribed message.
- Fixes (F) is the number of the correction keystrokes (such as delete, backspace, and cursor movement) and, modifier keys (shift, alt, control, etc.)
- Incorrect Fixed (IF) is the number of the incorrect keystrokes, which is noticed by the participant and corrected.

Error Rate measures can be calculated based on several definitions and as follows:

1) Error Rate of Minimum String Distance (ERMSD): Minimum String Distance is the minimum number of keystrokes needed to transform transcribed text into presented text [18]. ER_{MSD} is calculated by

$$\mathrm{ER_{MSD}} = (\mathrm{MSD(P,T)})/(\mathrm{MAX}(|\mathrm{P}|,|\mathrm{T}|))\times 100\% \quad (1)$$

2) Keystrokes per Character (KSPC): Another important error related metric is the keystrokes per character (KSPC), which it is value can be related either to many corrected errors, or to few uncorrected errors. In mathematical form, KSPC is obtained from

$$\mathrm{KSPC} = |\mathrm{InputStream}|/|\mathrm{T}| \quad (2)$$

3) Erroneous KeyStroke Rate (EKSR): The ratio of the total incorrect keystrokes to the presented characters is called EKS which calculated from

$$EKSR = (IF + INF)/P\times 100\% \quad (3)$$

4) Total Error Rate (TER): The ratio of total incorrect characters (fixed and non-fixed) to the total characters is called TER that is calculated from

$$TER = (IF + INF)/(C + F + INF)\times 100\% \quad (4)$$

IV. Materials and Methods

A. Dataset

An existing dataset [19] is adopted in this study. The original purpose of this dataset has been to investigate behavioral features of typing style to continuously authenticate smartphone users [20]. Data collection, which was previously reported in [21], was implemented using smartphone application named Hand Movement, Orientation, and Grasp (HMOG). The dataset was collected from 100 smartphone users (47 females and 53 males). Each user was asked to interact with smartphone with three types of tasks (reading session, writing session and mapping session). In this study, however, the writing session is of a particular interest.

The writing session data elements are used to extract the error rate as a performance metric while the corresponding keystroke dynamic metrics are processed to extract features and classify the user's fatigue status. The hold and digraph times are chosen to capture the typing pattern of the user throughout several writing sessions.

B. Features and Labels Preparation

1) Features vector preparation: The timing data of text entry keystroke are recorded with a high precision timer (with a resolution of 1μs) to ensure accurate and reliable data. Statistical representations of keystroke dynamics (i.e. the hold-time and the digraph time) are calculated to build a set of features. Some of these features such as HT mode, HT histogram area, diagraph mode, and diagraph histogram area are found less effective and thus eliminated.

Each one of the 100 participants has eight writing sessions to be completed at different periods. In each session, which has an average of 1193 taps and lasts for 11.6 minutes, the participant is invited to type three pre-specified statements using his/her smartphone's keyboard. In order to increase the number of records, each session is divided into 10 segments. So, for 100 participant with 8 session each multiplied by 10 segments yields 8000 records that form the size of the dataset used in this study. Next, the most effective four features are then selected to carry out this study; the hold-time mean, hold-time standard deviation, digraph mean, and digraph standard deviation.

2) Labels vector preparation: Many research papers and safety institutional reports agreed about the negative impact of fatigue on human performance [10], [22]. Text entry speed and accuracy is affected by mental fatigue [23]. The calculated metrics are divided into two sets; (i) a timing-

metrics set and (ii) an error-metric set. Several features of text entry are extracted from the collected data that are related to the typing performance measured in terms of speed and accuracy. Examples of performance metrics for a single participant are demonstrated in Figs. 2 – 4. In Fig. 2, the standard deviation of the hold time is changed between different sessions while the change in the number of backspace clicks is presented in Fig. 3. The backspace clicks is used as an indicator for fixed errors of the user typing. Similarly, the keystroke per character metric is presented in Fig. 4. A clear pattern's correlation between the results presented in these figures can be noticed which validates effectiveness of the selected features.

3) Dataset labeling: Based on the error metric, KSPC, a median threshold is chosen to label the feature's data for classification and training. Fig. 5 shows an example for the threshold-setting procedure that corresponds to one participant. As illustrated, the threshold line divides the writing sessions into two classes, alert and fatigued.

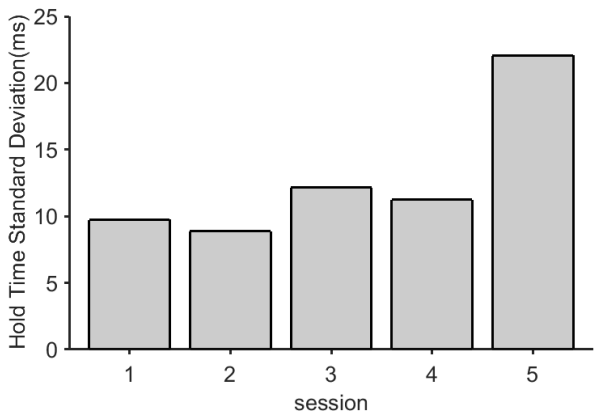

Figure 2. Example of hold-time standard deviation for different sessions

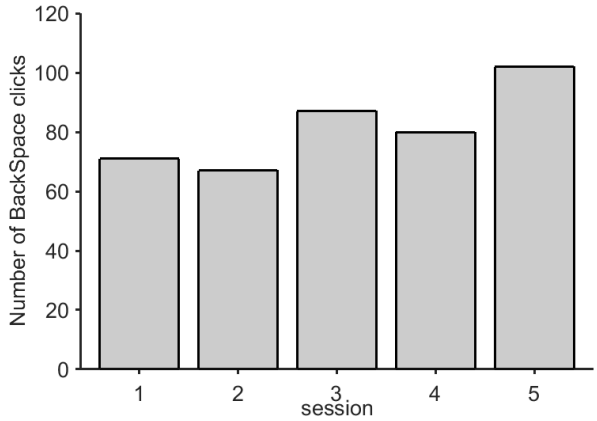

Figure 3. Example of backspace-clicks for different sessions

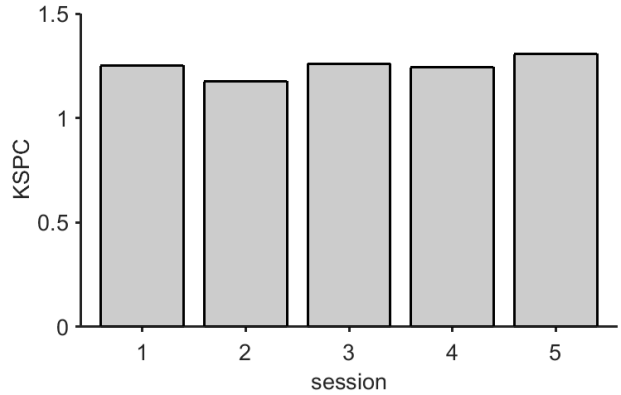

Figure 4. Example of KSPC for different sessions

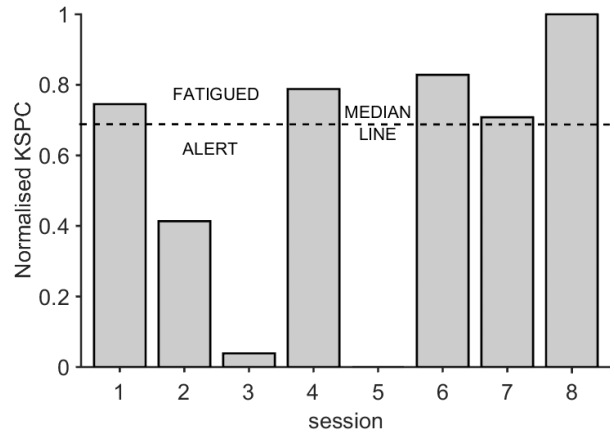

Figure 5. Threshold-setting for a 2-class labeling

V. IMPLEMENTED SYSTEM

A. System Description

The developed fatigue detection system comprises several stages as shown in the block diagram of Fig. 6. The collected smartphone's texting dataset is initially analyzed to calculate the keystroke dynamics as shown in the left branch of Fig. 6. This is achieved through filtering the timing data with the aim of reducing abnormality and preparing the collected data for the feature's extraction stage. Next, the fatigue-related features are then extracted as discussed earlier in Section IV-B. The selected features are then fed to a 2-class classifier, as illustrated. In the right branch of the diagram, the text-entry dataset is used to calculate the error (KSPC) that is used as an accuracy metric. Next, a median threshold is set depending on the calculated KSPC and the entire dataset is then labeled. The labeled dataset is used as a ground truth for the training stage of the classifier, as shown in Fig. 6. The classifier eventually classifies the users participated in this study depending on their fatigue/alertness status. In this study, two types of classifiers are implemented, ANN and SVM. These classifiers are described briefly as follows.

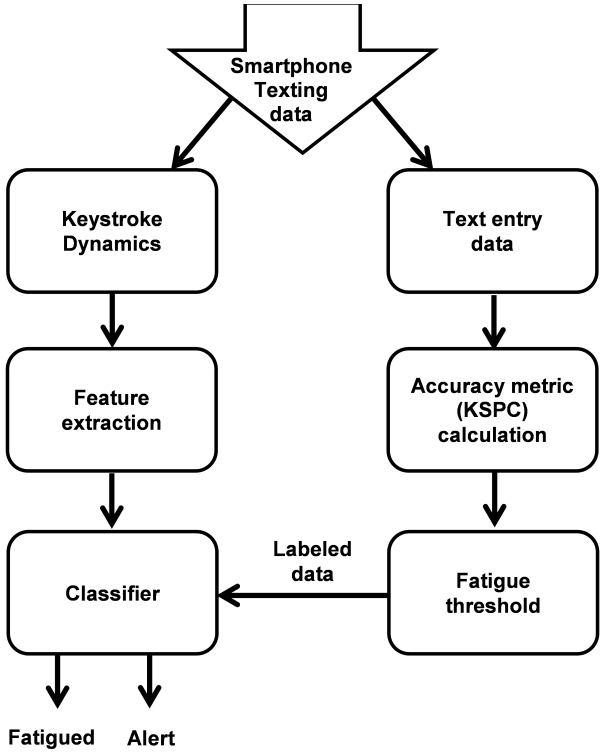

Figure 6. A block diagram for the developed fatigue detection system

B. *Artificial Neural Network*

ANN is built and trained with training set of 70% of the available dataset while the rest 30% records are used for validation and test phases. The proposed ANN is implemented with feed-forward ANN based on six neurons input layer and one hidden layer layers and one single output unit with a tangent-sigmoid transfer function. Moreover, many trials are adopted for ANN structure to improve the network performance such us changing the numbers of hidden layer and the nodes in each layer as well as changing in training algorithms and decision transfer function.

C. *Support Vector Machine*

SVM is a well-known classifier that has been successfully used in many applications [24]. In this study, the SVM is used to classify the users depending on their collected data into two classes (alert and fatigued). These classes are labeled to indicate the alert/fatigued status. Performance of the developed SVM classifier as compared to the ANN classifier performance is shown in the Receiver Operating Characteristic (ROC) graph of Fig. 7.

VI. Results and Conclusions

The developed ANN and SVM classifiers are trained and tested to assess the overall system performance. The obtained test results are summarized in Table I. It can be noticed that the SVM has a superior performance when compared to that of the ANN. These findings confirm the findings demonstrated in Fig. 7 since the area under the curve of the ROC for the SVM is larger than that of the ANN.

Unlike equivalent fatigue detection methods, which are mostly based on physiological and behavioral data, the proposed method does not need extra devices other than a smartphone. In addition, the proposed method is a non-intrusive method since it does not involve collection of any personal data, thus maintaining the users' privacy.

The suggested threshold setting and labeling procedure were found in agreement with the sleep-research findings reported in literature regarding using performance impairment as a fatigue sign/metric. The performance of the implemented classifiers showed SVM is superior to ANN in identifying fatigue/alertness conditions. SVM demonstrated accuracy of 88%, sensitivity of 84% and specificity of 100% when compared to the ground truth dataset.

These promising findings will facilitate the development of a low-cost and non-intrusive mobile instrument for fatigue/alertness detection. However, performance of the developed prototype can be further improved through utilizing timing data of user activities other than texting such as reading/scrolling, track-screen gestures and others. These improvements are currently part of the on-going research by the authors.

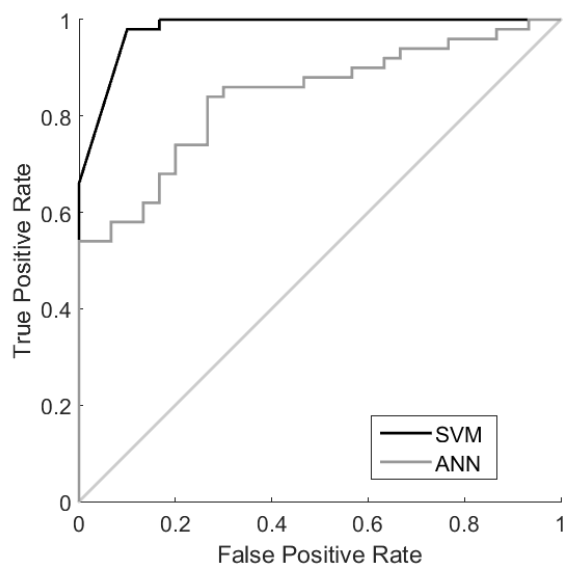

Figure 7. Example of ROC for one participant using SVM and ANN classifiers

TABLE I. Summary of ANN and SVM Performance Metrics

Performance metric	ANN	SVM
Accuracy	77.5%	88.8%
Sensitivity	79.6%	84.8%
Specificity	73.1%	100%

REFERENCES

[1] H. Abbood, W. Al-Nuaimy, A. Al-Ataby, S. A. Salem, and H. S. AlZubi, "Prediction of driver fatigue: Approaches and open challenges," in 2014 14th UK Workshop on Computational Intelligence (UKCI), Sept 2014, pp. 1–6.

[2] R. O. Phillips, "What Is Fatigue and How Does It Affect the Safety Performance of Human Transport Operators? Fatigue in Transport Report I," TØI Report, no. 1351/2014, 2014.

[3] V. J. Gawron, J. French, and D. Funke, "An overview of fatigue," in Stress, workload, and fatigue. Lawrence Erlbaum Associates Publishers, 2001, pp. 581–595.

[4] H. Al-Libawy, A. Al-Ataby, W. Al-Nuaimy, M. A. Al-Taee, and H. S. AlZu'bi, "Estimation of driver alertness using low-cost wearable devices," in Applied Electrical Engineering and Computing Technologies (AEECT*),* 2015 IEEE Jordan Conference on, Nov 2015, pp. 1–5.

[5] H. Allibawy, A. Al-Ataby, W. Al-Nuaimy, and M. A. Al-Taee, "HRV-based operator fatigue analysis and classification using wearable sensors," Proc. 13th International Multi-Conference on Systems, Signals & Devices (SSD), Germany, 2016, pp. 268 – 273.

[6] Z. J. Muhsin, A. Al-Taee, M. A. Al-Taee, W. Al-Nuaimy, A. Al-Ataby, 2016. Mobile workflow management system based on the Internet of Things, Proc. 13th International Multi-Conference on Systems, Signals & Devices (SSD), Germany, 2016, pp. 436-441.

[7] A. M. Al-Taee, A. Al-Taee, Z. J. Muhsin, M. A. Al-Taee, W. Al-Nuaimy, "Towards developing online compliance index for self-monitoring of blood glucose in diabetes management," Proc. 9th International Conference on Developments in eSystems Engineering (DeSE '2016), Liverpool & Leeds, UK, 2016.

[8] A. M. Al-Taee, M. A. Al-Taee, W. Al-Nuaimy, Z. J. Muhsin, H. AlZu'bi, "Smart bolus estimation taking into account the amount of insulin on board," Proc. IEEE Conf. on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), Liverpool, UK, 2015 pp. 1051-1056.

[9] M. A. Al-Taee, S. N. Abood, W. Al-Nuaimy, A. M. Al-Taee, "Blood-glucose pattern mining algorithm for decision support in diabetes management," Proc. 14th UK Workshop on Computational Intelligence (UKCI). Bradford, UK, 2014, pp. 1-7.

[10] N. A. Capela, E. D. Lemaire, and N. Baddour, "Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients," PloS one, vol. 10 (4), p. e0124414, 2015.

[11] X. Su, H. Tong, and P. Ji, "Activity recognition with smartphone sensors," Tsinghua Science and Technology, vol. 19 (3), 2014, pp. 235– 249.

[12] R. W. Soukoreff and I. S. MacKenzie, "Metrics for text entry research: an evaluation of msd and kspc, and a new unified error metric," in Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2003, pp. 113–120.

[13] A. S. Arif and W. Stuerzlinger, "Analysis of text entry performance metrics," in Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference, 2009, pp. 100–105.

[14] R. W. Soukoreff and I. S. MacKenzie, "Measuring errors in text entry tasks: an application of the levenshtein string distance statistic," in CHI'01 extended abstracts on Human factors in computing systems. ACM, 2001, pp. 319–320.

[15] P. S. Teh, A. B. J. Teoh, and S. Yue, "A survey of keystroke dynamics biometrics," The Scientific World Journal, vol. 2013, Article 408280, Available on: http://dx.doi.org/ 10.1155/ 2013/408280.

[16] W. He, "Evaluating differences between keystroke datasets collected under lab and field conditions," Ph.D. dissertation, Carnegie Mellon University Pittsburgh, PA, 2015.

[17] M. Koivisto and A. Urbaczewski, "Accuracy metrics in mobile text entry," in Proc. 1st IASTED Conf. on Human-Computer Interaction, 2005.

[18] F. J. Damerau, "A technique for computer detection and correction of spelling errors," Communications of the ACM, vol. 7, no. 3, pp. 171–176, 1964.

[19] H-MOG Dataset, Available on: http://www.cs.wm.edu/ ~qyang/hmog.html, (Accessed 02 July 2016).

[20] Z. Sitova, J. Sedenka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. Balagani, "Hmog: New behavioral biometric features for continuous authentication of smartphone users," 2015.

[21] Q. Yang, G. Peng, D. T. Nguyen, X. Qi, G. Zhou, Z. Sitov´a, P. Gasti, and K. S. Balagani, "A multimodal data set for evaluating continuous authentication performance in smartphones," in Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems. ACM, 2014, pp. 358–359.

[22] N. Lamond and D. Dawson, "Quantifying the performance impairment associated with fatigue," Journal of sleep research, vol. 8(4), 1999, pp. 255–262.

[23] A. Pimenta, D. Carneiro, P. Novais, and J. Neves, "Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns," J. Hybrid Artificial Intelligent Systems, Springer, 2013, pp. 222–231.

[24] A. Y. Al-Hyari, A. M. Al-Taee, M. A. Al-Taee, "Diagnosis and classification of chronic renal failure utilizing intelligent data mining classifiers," Int. J. Information Technology and Web Engineering (IJITWE), 9(4), pp. 1-13, 2014, doi: 10.4018/ijitwe.2014100101.