Neural Comput & Applic
DOI 10.1007/s00521-014-1790-y

ORIGINAL ARTICLE

Hybrid affective computing—keyboard, mouse and touch screen: from review to experiment

Kaveh Bakhtiyari · Mona Taghavi · Hafizah Husain

Received: 9 April 2014 / Accepted: 8 December 2014

Abstract Emotions play an important role in human interactions. They can be integrated into the computer system to make human–computer interaction (HCI) more effective. Affective computing is an innovative computational modeling and detecting user's emotions to optimize system responses in HCI. However, there is a trade-off between recognition accuracy and real-time performance in some of the methods such as processing the facial expressions, human voice and body gestures. Other methods lack efficiency and usability in real-world applications such as natural language processing and electroencephalography signals. To accomplish a reliable, usable and high-performance system, this paper proposes an intelligent hybrid approach to recognize users' emotions by using easily accessible and low computational cost input devices including keyboard, mouse (touch pad: single touch) and touch screen display (single touch). Using the proposed approach, the system is developed and trained in a supervised mode by artificial neural network and support vector machine (SVM) techniques. The result shows an increase in accuracy of 6 % (93.20 %) by SVM in comparison with the currently existing methods. It is a significant contribution to show new directions of future research in emotion recognition, user modeling and emotional intelligence.

Keywords Affective computing · Human emotion recognition · Keyboard keystroke dynamics · Mouse touch pad movement · Touch screen monitor · Human–computer interaction (HCI)

K. Bakhtiyari (✉)
Interactive Systems, Department of Computer and Cognitive Science, Faculty of Engineering, University of Duisburg-Essen, 47048 Duisburg, North Rhine-Westphalia, Germany
e-mail: academic@bakhtiyari.com

K. Bakhtiyari · M. Taghavi · H. Husain
Department of Electrical, Electronics and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, UKM (The National University of Malaysia), 43600 Bangi, Selangor Darul Ehsan, Malaysia
e-mail: mona@siswa.ukm.edu.my

H. Husain
e-mail: hafizah@eng.ukm.my

1 Introduction

Human interaction is an important role in human communication. It builds trust and exchanges beliefs. These interactions can be categorized as a verbal or nonverbal communication. Human emotion is a type of nonverbal message, which plays an important and effective role in communications. People understand each other's emotions in their interactions, and it leads to a better and more reliable communication.

Nowadays, people spend a lot of time with digital gadgets such as personal computers, PDAs, tablets and smart phones. Intelligent systems are moving forward by providing means of communication among their users by employing various methods and technologies. They try to understand their users' needs, therefore, to personalize their interface. Recognition of human emotions is a step into the future of artificial intelligence to have computers behaving more similar to human.

Human emotion recognition systems fall into various different categories from gaming to business applications. Emotional intelligent systems can respond to the users according to their emotions and build a connection between computer and users more naturally [1].

Users often talk about their computers, and they describe the interface as the system underlying it, which is almost consistent among the majority of users [2]. In order to make people-friendly technologies, interfaces are modeled based on the people who interact with it. In the development of application interfaces, it is necessary to incorporate all aspects of human behavior, including cultural and social competence and awareness in the design considerations [3].

Generally, a transmitted message consists of two major channels, which are sending explicit and implicit messages. Explicit messages are about the message owner features. Implicit messages can be about anything, and not even special. Implicit messages are not very well known, which require a lot of efforts to comprehend. The emotions of the speaker are included in the second type of channel in implicit messages [4]. Emotions are discussed by three parameters. The first parameter is *Arousal*, which shows the energy of feeling. In the literature, they are classified with different names as emotions. Happiness, sadness, joy, etc. are examples of arousals. The second parameter is *Valence*. Valence presents whether the feeling is pleasure (positive) or displeasure (negative) in case of the energy. And the third parameter is Dominance. Dominance shows the strength of the emotion, and it explains how strong an emotion is. Most of the current emotional-aware systems consist of only arousal and valence parameters. Integration of dominance is useful if and only if a high accurate recognition of arousal is achieved.

Affective computing is a new research area, which studies the recognition, interpretation, process and simulation of human affects [5]. Due to the novelty of this area, it suffers from serious challenges such as: (1) performing in real time, (2) reliable recognition accuracy and (3) applicable on most of available computers. This paper presents an accessible, achievable and implementable hybrid method for a large group of available computer systems. This methodology recognizes human emotions by using common input devices such as keyboard, mouse (touch pad) and touch screen displays. These devices are available on most of currently available personal computers.

1.1 What is emotion?

Figure 1 illustrates the states of emotions and their persistency in time. Emotions that last only few seconds are *Expressions*; recognizing the expressions is not useful, and they are very complicated because of their short life span. The second state is very similar to expressions, but it may also last alongside with expressions. This type is called *Autonomic Changes*. It takes a few seconds; then, a little later, it does not exist anymore. The third state of the emotions that last from seconds to minutes is *Attitude*. Attitudes recognition by the system or human might be useful. The fourth state is *Self-Reported (full blown) emotions*, which may last from minutes to hours. Most researchers working in affective computing are concentrating on this state of emotion. This emotional state is guaranteed to be fixed for a time to be processed and recognized. Therefore, the user's computer system has enough time to respond properly to these emotions. The next type of the emotional state is *Mood*. Mood is a kind of emotion between hours and months. The sixth state is *Emotional Disorders*, which lasts from weeks to years. Finally, the seventh is *Traits,* which may last for a lifetime, and it can be considered as a part of human characteristics. Moods and emotional disorders can be recognized by further processing of the Self-Reported emotions [4].

Some mental states are placed in the different emotional states above. They may exist as expressions or even as traits. For example, happiness can be considered as a Self-Reported (full blown) emotion, mood or even a trait [6].

1.2 Classification of emotions

In psychological studies, emotions are classified in various groups. Some researchers classified the emotions based on their own requirements on the research. In Sanskrit, a group of nine emotions is selected as the *Basic Emotions* as below [7]:

- Sexual passion, love, delight
- Amusement, laughter, humor, mirth
- Sorrow

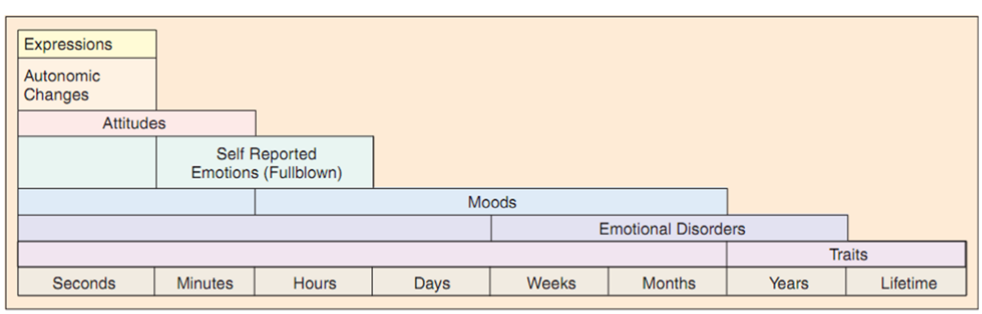

Fig. 1 States of emotion categories based on the time

- Anger
- Fear, terror
- Perseverance, energy, dynamic energy, heroism
- Disgust, disillusion
- Amusement, wonder, astonishment, amazement
- Serenity, calm

Another classification of emotions is done by Plutchik, which is known as Plutchik's *emotion wheel* [4]. In this classification, eight emotions are described in a circular way that each emotion is close by properties of the next emotion by an angle. These emotions are illustrated in Fig. 2.

Plutchik emotions list is the most common used classification in affective computing. Cowie et al. presented a table of emotional words from Whissell and Plutchik to represent each emotion in three dimensions of Activation, Evaluation and Angle [2]. Different emotions are classified in various rates, and the resulted value would be matched to the rates to detect the proper emotion related to the context. Angle is called emotional orientation. For example, the value for Acceptance is considered 0, Disgust is 180, Apathetic is 90 and Curious is 270. The other emotions between these emotions on Plutchik emotion wheel have a range between them. This measurement is based on a circular degree from 0 to 359. Activation is another important feature, which shows the possible emotion states. Activation value has been determined by Whissell. There are some differences in Plutchik and Whissell methods of representing the emotions. For instance, Fear and Anger are two opposite emotions in Plutchik's wheel, but they are close in Whissell's method. Most research has used Whissell activation for their computations [2].

Emotions are universal expressions. In a comparative cultural study, people were asked to identify which emotion could be seen in standardized photographs of faces? It was shown that people all over the world can accurately assign these aforementioned emotions using the facial expression; thus, the question could be answered by universal emotions [8].

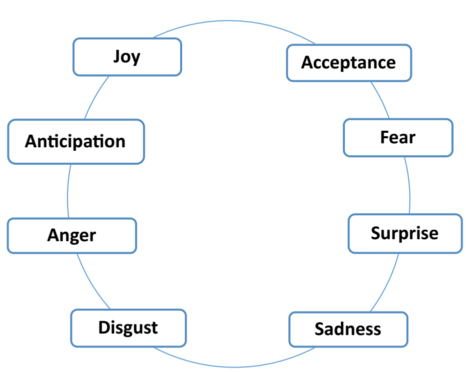

Fig. 2 Plutchik's "Emotion Wheel" [4]

2 Affective computing methods and technologies

Affective computing has been a challenging and attractive research area for a long time. Researchers have tried various methods and techniques to achieve their goals of recognizing the proper emotion of the computer users. Some of the major methods and technologies, which recently were used, are listed below:

- Facial expression recognition
- Body gesture recognition
- Natural language processing (NLP)
- Voice recognition
- Electroencephalography (EEG) signal processing
- Physical human–computer interaction:
    - Keyboard
    - Mouse (touch pad)
    - Touch screen monitor (single touch)

Facial expression and body gesture recognition are mostly based on the image processing techniques. They work by capturing the facial image of the user and processing the video of the body movements to recognize the emotion. These two methods are the most common methodologies in affective computing. NLP and voice recognition work on the language patterns. They process the pattern of the user's talks by analyzing the words, definitions and intonations. Electroencephalography (EEG) signal processing is the most recent method. This device was previously used in medical experiments. The other method is through normal input devices like the keyboard, mouse (touch pad) and touch screen, which are ubiquitous.

2.1 Facial expressions recognition

Facial expressions are the reflex of different emotions on the human face; actually, it is the change of muscles position in the face, eyes or skin to show an emotion [9]. There are basically six emotions and about 17 sub-emotions, which can be classified in facial expressions [10]. Those six emotions are *Joy*, *Surprise*, *Fear*, *Anger*, *Disgust* and *Sadness*, and the 17 sub-emotions are listed as *Anger*, *Concentration*, *Confusion*, *Contempt*, *Desire*, *Disgust*, *Excitement*, *Empathy*, *Fear*, *Flirt*, *Frustration*, *Glaring*, *Happiness*, *Sadness*, *Snarl*, *Surprise* and *Love*.

Facial expressions are considered as nonverbal communication. It sends the emotion to the other side of the communication channel without expressing a single word or phrase. In some countries like USA, the facial expressions are used in the sign language [11]. Facial expressions are different from one country to another. In Asian countries, the intensity of facial expressions is less than the other countries, because among some Asian cultures, it is rude to show some specific emotions on the face. Therefore, showing some negative emotions may cause a disharmony in the society [7].

For more than a decade, researchers in computer science have tried to recognize emotions by processing the facial expressions. This method is based on image processing. It captures a facial image. Then, it extracts the feature points and analyzes the location of each feature point with the other points. At the end, it can recognize the proper emotion from the facial image [12–14]. Figure 3 demonstrates the feature points on two volunteers for facial expression recognition.

As image processing techniques are time-consuming with high computational time, decreasing the number of feature points on a face may decrease the computational time. However, the accuracy of the system remains reliable. Some researchers used human visual cortex (HVC) and radial encoding to improve the performance [15]. Konar et al. [16] used type-2 fuzzy model on facial expressions to recognize emotions. He partially succeeded in multi-level recognition of emotions (dominance parameter), and he gained 88.66 % in recognition accuracy. Kao and Fahn [17] used a combination of several machine learning methods as ABA and CART on facial expression, and they achieved a high accuracy of 90 %. Ilbeygi and Shah-Hosseini (2012) used fuzzy inference system (FIS) on facial expressions with the accuracy of 93.96 %. In addition, Genetic Algorithm was applied to tune the parameters of the membership functions [18]. 95.24 % of accuracy on the facial expression datasets of JAFFE and MUG is achieved in 2013 [19]. They used the cloud computing to perform their research. This result was based on the analysis on popular databases. However, the images taken in a real situation would be different and troublesome due to the image noises.

Facial expression recognition has some advantages that they are known as the positive points of this method. These positive points are as follows:

- It is based on image processing, which is a supervised method.
- A normal camera can be used for this method.
- It can be integrated with body gesture recognition.
- Extra (explicit) information such as the estimation of the user's age and gender can be extracted by image processing.

Fig. 3 Feature points on facial expressions recognition are shown in white dots

As the image processing techniques are the positive points of this method, itself, it is also a negative point for facial expression recognition. The process complexity of the facial expression recognition cannot be reduced to a simple algorithm with low computational cost [20]. The available challenges in facial expressions recognition are listed below. These challenges make the system far from real-time processing and anytime availability:

- Image processing is time- and resource-consuming.
- A camera is required, which may not be available in all cases.
- The user should be facing toward the camera, and the head direction changes may distract the recognition process.
- Multiple faces in captured picture by camera may confuse the system in the procedure.
- Noises and external objects in the picture may interrupt the recognition process.
- Privacy issues by using video camera exist for many users.

Besides the above problems in facial expressions, two emotions may have similar facial expressions in different cultures. Therefore, it would be difficult to recognize the proper emotion. A hybrid method of recognition by using facial expression recognition in parallel with other methods may come as a solution to overcome this problem.

2.2 Body gesture recognition

Body gesture is another nonverbal communication among people. It exchanges messages without textual or vocal phrases. Body gesture is a visible and visual message, which is mostly conducted by hands and head. In some gestures, the whole body can be used as a sign. In some cultures, body and hand gestures are used in parallel with the speech to clarify and emphasize the concept of the speech [21]. They are also known as a mean to transfer the emotions while speaking [22].

History of body gesture goes back to more than 500 years ago. There are lots of people who worked on the gesture to analyze the language structures or even to describe human personality through the gestures. John Bulwer in 1644 [23] discussed about body gestures, their importance and their usage in the daily speaking. Later on, in 1832, Andrea De Jorio [24] elaborated and extended the definition of gesture expressions. Since then, this research area has been continued, and researchers are still working on the different important aspects of gestures on human life. For instance, we can name David McNeill [25] or even Susan Goldin-Meadow [26] who has worked in this area.

Body gesture recognition can be done in online and offline modes, and the pictures can be rendered and processed in 3D and 2D (appearance based). The affective computing extracts the features of body gesture to recognize the emotions of the user. Gunes et al. [27] even gained an accuracy of 81 % in human emotions recognition by analyzing the body gestures, and some other researchers like Glowinski et al. [28] only tried to propose a new solution for processing the body gesture toward the recognition of emotions. Chen et al. [29] tried a hybrid model of body gesture and facial expressions recognition. In their initial test, they evaluated each body gestures and facial expressions separately, and they could achieve an accuracy of <70 %. And in the hybrid model, they improved their initial result up to 72 %.

Image processing is used in body gesture recognition similar to facial expression recognition. So, the advantages and disadvantages of this technique are limited to image processing limitations, and they are similar to the limitations of facial expression recognition. Beside good and reliable accuracy of this method, high computational time and resource-consuming are the main weaknesses of this technique.

Privacy is an important issue that users do not feel comfortable to be recorded by video cameras.

2.3 Natural language processing (NLP)

The other way of emotional exchange among people in social life is using words. Each adjective, verb or generally a word can represent emotional states in a sentence. An obvious example of this natural state is recognition of a person's emotion while somebody is reading a letter or a short message (SMS).

Ortony, Clore and Collins (OCC) model [30] with 22 emotion rules was also used for human emotion recognition. This model presents a knowledge base (KB) of words and emotions, which are customized for each person. Li et al. [31] enhanced and improved the OCC model from 22 rules into 16 rules. They also used five-factor model (FFM) to analyze the emotional rules on each dimension. At the end, they tested their own system by two experiments, and they gained their best result in Anger for 74.1 % and the worst in Pity for 58.1 %.

Social networks were also concerned in some research. For instance, Twitter has 340 million tweets per day. In 2013, Spanish Twitter corpus was evaluated with the accuracy of 65 % on emotions recognition by applying multinomial naïve bayes (MNB) machine learning method on *n*-gram features [32]. In this research, unigram presented the best results on classification.

Calvo and MacKim [33] developed two different computational models based on NLP for emotion recognition. They used four datasets with an emotional thesaurus and a bag of words. LSA, PLSA and NMF dimensionality reduction techniques are employed to evaluate the recognition. Their best result was related to Anger/Disgust and Joy with the accuracy of 77.3 % by using NMF-based categorical classification (CNMF) [33]. Their achieved performance is near to Li et al. (2008) with 75.9 on distress.

NLP in affective computing is improving, but still it is far from being used in commercial applications. So far, only a few types of communication are based on the text, and this method is only available on few resources and media. Moreover, language differences cause different emotional patterns, and NLP should be done separately on different languages and cultures [7].

2.4 Speech signal processing

Voice/Sound/Speech processing is the science of analyzing the tones and the spectrum of the sound to extract a meaningful concept. The voice recognition method is based on the signal processing techniques. A very common usage of voice recognition is converting human voice to the text, which can be found in recent computers, handheld devices and mobile phones. Speech processing has stronger literature for user identification. To date, in speech signal processing, the average accuracy of user identification/authentication is higher than emotion recognition [34].

Emotions have a direct influence on the human voice. This may result in differences in voice tones and vibrations. A natural example of this nonverbal message is recognition of emotion while you are talking on the phone [35, 36]. These changes can be marked in a sound spectrum as identification. This research area goes back to *query by humming* [37], which basically works on searching through sounds. Then, by improving those methods, we are able to mark and identify these emotional changes inside the voice. Generating identification of emotional states in the sound is called fingerprint. The fingerprint is nothing more than marked top frequencies in various and different frequency rates in a spectrum [38]. After that, these fingerprints would be matched with a collected series in database or knowledge base to retrieve an equivalent emotion.

Despite all the above efforts, the recognition accuracy of this method is still low, and it is unreliable to be used independently in a system. So, it is preferred to be used in a hybrid with the other methods. Amarakeerthi et al. [39] combined it with postures and body gestures; Kleckova and Pittermann et al. [35, 40] integrated with NLP methods. Wang et al. and Hunag [36, 41] integrated it with visual and facial recognition. Although emotion recognition is possible through the sound, this technique can be used to identify the gender [42]. The registered patent in 2014 in USA by Krishnan and Fernandez only shows the accuracy of 40–50 % in emotion recognition by analyzing the human speech. They used statistical and MFCC features to compare the speech and reference sample in order to recognize the proper emotion [43].

2.5 Electroencephalography (EEG) signal processing

Electroencephalography (EEG) records electrical activities and changes of neurons along the scalp. It works by measuring the voltage of ionic current flows in the neurons of the brain. This device was previously used in neurology and medical purposes, but today as its cost has decreased, computer scientists use it in their own applications. There are some laboratory-based computer games and applications, which work by using EEG. Figure 4 shows an installed EEG on the scalp of a woman.

EEG works with a high accuracy as it works directly with the brain activity. Liu et al. [44] in Singapore used machine learning and collected EEG data (signals) with the labeled emotions for training. Later, these results were compared with the valence and arousal of real emotions in the detection. This is the general procedure of machine learning methods, which works based on training and testing (application) sessions iteratively. Schaaff [45] has also tried this device in affective computing and reached 66 % of accuracy. Then, Guangying et al. [46] used support vector machine (SVM) to improve the performance of the system with the reported recognition rate of more than 83.33 %.

Most of the reported performances on emotion recognitions by EEG are not competitive with the other available methods. However, in 2013, 91.33 % of accuracy by EEG is achieved by using four frequency bands, namely alpha, beta, gamma and alpha to gamma. Probabilistic neural network (PNN) and K-nearest neighbor (KNN) are used in their research, and their highest accuracy was made by KNN [47].

Despite the reliable results of EEG signal processing, this device is not available and cheap to be easy to access.

Fig. 4 A simple installed electroencephalography (EEG)

Also, installation and maintenance of this device require technical experts and trainings. This device is only accessible and usable in laboratory/research level. It might be a long time before it is used as a common device among computer users and to be sold as an accessory.

2.6 Common input devices

Every day, people are using some regular input and output devices to interact with digital systems. Mice (touch pads), keyboards and touch screen monitors are the common input devices. At least one of these devices is available and accessible on all computers and digital systems [48–50].

Monrose et al. authenticated the computer users by keystrokes dynamics on keyboard in 2000. They tested on 17 computer programmers. The system was successful to identify all users by their keystroke patterns while typing a specific word or phrase with the accuracy of 92.14 %. However, there was a challenge in his research. At the time of training and testing the system for authentication process, the users should be in the same emotional state. Otherwise, the keystroke pattern would be changed; therefore, the computer cannot identify the user properly [51, 52]. Meanwhile, they implied that human emotions reflect on the keystroke dynamic patterns.

Schuller et al. [48] from Technical University of Munich in 2004 worked on emotion recognition by analyzing the mouse movement patterns and clicks on the screen. They used the SVM to train the system and achieved a reliable accuracy of emotion recognition on four major emotions. They got a recognition accuracy range between 77.7 and 90.2 % for mouse interaction and the range of 43.3 and 92.7 % for touch screen interactions. They also performed a hybrid experiment of mouse and touch screen interactions.

Milanova et al. [49] tried to increase the reliability of the facial recognition by integrating keyboard keystrokes analysis. Their research showed an average of 8 % improvement by using a combination of methods rather than a single facial expression at that time. Epp et al. [53] used keyboard keystroke dynamics to classify emotions. They achieved the highest accuracy of 87.8 % for Sadness, and the least 77.4 % for Neutral or Relaxation. They also used Kappa statistics to measure the membership value of the emotions in classification. They used C4.5 supervised machine learning method in Weka software [54] for classification.

Common input devices are available and easy to access to be used for human emotions recognition. To the best of our knowledge, mostly only one input device is used to be processed and analyzed; however, a combination of devices is preferable. The main weakness of using input devices is having different patterns of usage by various users. For example, users' keystroke dynamics on keyboard are different, and the level of proficiency in computer alters the results of the analysis. In such case, the result of using input devices cannot be extended to other users without consideration of users' similarities.

2.7 Critiques

As discussed earlier, there are still unanswered problems and open challenges in this research area. Some of those challenges are covered in this research as we walked through and analyzed the possibilities.

Even though facial expression recognition gained a high accuracy in emotions recognition, in real-time processing, still it is inferior to the other methods. Because image processing techniques are time- and resource-consuming, NLP and common input devices can be used for real-time applications. However the reported recognition results of NLP and common input devices in recent research literature are not satisfying.

Some technologies such as EEG machines are expensive and not easily available and usable in daily life. The other methods of affective computing which use microphone, camera and other input devices are much cheaper and more available. But because of the security and privacy issues, many users may not allow usage of microphones and cameras. This issue limits the number of developed applications for facial expression recognition, body gestures recognition and voice processing.

Even though the ability of emotion recognition by common input devices has been proved, but the low accuracy still keeps the door and questions open in the affective computing research area. Table 1 shows the positive and negative features of various methodologies in affective computing to detect human emotions. In this table, the best achieved recognition accuracies are also listed.

2.8 Problem statement

Affective computing can be applied in commercial and daily software applications, when the system can achieve a high reliable accuracy and low false-positive in emotion recognition. Affective computing has been utilized by various methods and techniques to achieve this goal, but still is far away. There are some challenges from different perspectives, which make it an open research area to work on [4, 59, 60]. The first challenge is to achieve a higher accuracy in emotion recognition with a reliable precision (relatively low false-positive rate) [12, 27, 56]. The available techniques are not reliable and accurate enough to be employed in real-world applications. Every day, new techniques and methods are being introduced, and computer systems are becoming faster and smarter. Therefore,

Table 1 Pros and cons of different methodologies in affective computing and their highest achieved accuracy

Method	Accuracy (%)	Pros	Cons	References
Facial expression	95.24	Supervised method Cheap equipment Extra information	Time- and resource-consuming Noise Image processing problems Privacy issues	[15, 20, 29, 55, 56]
EEG	91.33	It can be extended for real-time processing	Expensive device Low recognition accuracy on dominance emotion	[44–47]
Body gesture	81	Supervised method Cheap equipment Extra information	Time- and resource-consuming Noise Image processing problems Privacy issues	[27, 29]
Voice recognition	78.64	Accurate Integrated into interactive user interfaces	Cultural and language differences Time- and resource-consuming	[34–36, 39, 41–43]
NLP	77.30	Easy implementation	Not accurate enough Cultural and language differences Not being real time	[31–33, 57, 58]

we can apply new methods and hypotheses to gain better results with a higher performance. The second challenge is the real-time processing [16, 41, 51]. Image and signal processing are the most common techniques used in emotion recognition. However, both techniques are time and computational resource intensive. It is very important to be able to recognize the user's emotion in real-time online. The third challenge is using the available and cheap hardware for the recognition process [49, 52, 53]. For instance, the EEG devices are not easily available, not-portable, difficult to install and expensive. The desired solution here is to work on the ability of identifying emotions by available, easy to carry and cheap devices such as normal input devices like mouse (touch pad), keyboard and touch screen displays.

2.9 Trust

The recent problem, which was introduced by Simon Sinek at a TEDx presentation in Maastricht, Netherlands [61], is the lack of direct interaction among the people in a society by using the technology and electronic media. This phenomenon made the communications faster than before, but the level of trust in human life has been decreased. Trust is a human interaction. It is made among the people who believe what the others believe. Meanwhile, electronic communication degraded the happiness and passion of communication. This split can be healed by importing the natural human emotions into the digital communication [60].

This research proposes a solution for human emotions recognition to address the three mentioned problems. We applied a hybrid methodology on analysis of the users input using common input devices such as keyboard keystroke dynamics, mouse (touch pad) movements, and touch screen monitor interactions. Combining the results of analysis of three devices will provide a higher accuracy in emotions recognition as these devices are available on most of computer systems. Fast learning techniques for data classification are chosen to provide faster recognition and to be closer to real-time processing.

3 Methodology

The methodology is based on prototype software, which records the user's interaction data from mouse (touch pad), keyboard and touch screen interactions. This methodology is known as experience sampling methodology (ESM) [62]. A prototype application was designed and developed to collect the required data from users' interactions. This software was installed on the volunteers' computers for a specific period of time (1 month) to process and analyze users' emotions.

In this study, the universal emotions by Paul Ekman [56] have been selected. Then, the dataset was minimized into four emotions. This action helps to minimize the data scattering in the recorded dataset and also to compare the results with the other scholars. These four emotions are as follows:

- Neutral (includes emotion of happiness and as perceived normal mood)
- Fright (afraid) (includes helplessness, confusion and surprise)
- Sadness (includes primarily sadness, anger and resentment)
- Nervousness (including nervous, fatigue and lightheadedness)

The keyboard keystroke dynamics, mouse (touch pad) movements and touch screen interactions of 50 users were collected. Every 4 hours, users were asked to enter their current emotion (Self-Reported emotion). This procedure continued for 1 month; then, the collected data were used in RapidMiner for classification.

For evaluation of the mouse (touch pad) and touch screen interaction, the methodology presented by Schuller [48] has been used. For data collection, all the mouse movements and mouse (touch pad) keystrokes were collected; meanwhile, in the similar research, only a limited number of the features were analyzed. Also, evaluation method of the touch screen interactions is retrieved from Schuller et al. keyboard features that were also presented by Monrose and Rubin [51] for authentication purposes; however, the emotions were the weakness of their research. This weakness arises from the emotions interrupting the authentication process during keyboard keystroke dynamics. In this research, the keyboard keystroke dynamics features used to analyze the emotions are based on their study.

These 50 users were selected with various cultural backgrounds. Then, the prototype application was installed on each of the user's personal computer. Prototype application recorded the interactions for 1 month and prompting the users to enter their own proper emotions with defining the level of each emotion. These users were mostly settled physically in Malaysia, Germany and Iran.

One of the most important processes of data preparation is data cleaning to remove all redundant and inaccurate/incomplete entries from data. As these data are recorded by a prototype application, each entry is checked at the time of recording to avoid the representation of incomplete data. No data collection in this research has been conducted manually.

3.1 Keyboard

Keystroke dynamics are habitual rhythmic patterns of typing. These features are used in biometrics for identifications for more than a decade. Representation shows the input values as the words. When the user is typing, he is actually representing his identity. The next step is features extraction where the system extracts and defines the features as a fingerprint and records them in a database. The last section is classification that matches the extracted features of a new user with the existing features in the database to identify him/her. This research has used the similar method, but there are differences to identify the emotions instead of the users.

3.1.1 Keystroke dynamics features

There are three major features in keystroke dynamics as below:

- Key down-to-down
- Key down-to-up
- Key up-to-down

The above features have worked well on users' identification in a neutral emotion. The first feature, the key down-to-down feature measures the time between subsequent key presses. This feature has two hidden parameters as duration and latency. These parameters are basically the next features of keystrokes. Key down-to-up is the exact time between pushing and releasing a button. This item can also be considered as duration. Duration is the spent time for one character. The last feature is key up-to-down that it is also called latency. Latency is the wasted time between typing two characters. Research showed that the latest two features are 10 times more in a novice typing in comparison with a professional and expert user [51].

The keystroke features (KFs) were selected from the timing differences of single keystrokes, digraphs (two-letter combinations) and trigraphs (three-letter combinations) [53]. For each feature, the mean and standard deviation are calculated, because during a sample period, the user might enter the same sequence of keys more than once (e.g., entering 'th' twice during the sampling period). The following 15 features are defined for the keyboard keystroke dynamics:

- *KF1* The duration between first and second down keys of the digraphs.
- *KF2* The duration of the first key of the digraphs.
- *KF3* Duration between first key up and next key down of the digraphs.
- *KF4* The duration of the second key of the digraphs.
- *KF5* The duration of the digraphs from first key down to last key up.
- *KF6* The number of key events that were part of the graph.
- *KF7* The duration between first and second down keys of the trigraphs.
- *KF8* The duration of the first key of the trigraphs.
- *KF9* Duration between first key up and next key down of trigraphs.

- *KF10* The duration between second and third down keys of the trigraphs.
- *KF11* The duration of the second key of the trigraphs.
- *KF12* Duration between second key up and next key down of trigraphs.
- *KF13* The duration of the third key of the trigraphs.
- *KF14* The duration of the trigraphs from first key down to last key up.
- *KF15* The number of key events that were part of the graph.

3.1.2 Keystroke dynamic framework

As a standard framework, there are twenty (20) sets of characters, which are defined as a fingerprint for keystroke dynamics. A graph of duration and latency for each of these 20 sets represents a unique identity that was used for user identification. This fingerprint is applied to match with the user's emotions instead of user's authentication. These twenty sets are as follows [51]:

th	er	is	or
he	an	at	hi
nd	ng	on	the
re	me	es	ing
in	we	ay	are

The above sets are chosen in English language as they have been occurred the most in an English context. Users in different emotional states type the words by different durations and latencies, and the above sets are chosen to be recorded as the user's fingerprint.

Figure 5 shows a sample of recorded keystroke dynamics of a user with a neutral emotion. This figure is considered as the user's fingerprint in a neutral emotional state.

3.2 Mouse (touch pad: single touch)

It seems reasonable to divide the mouse (touch pad: single touch) movements into two different groups. The first group is the movement of the mouse without using the left mouse (touch pad) button pressed. The second group is where the mouse button is pressed. An attempt is made to dwell mainly on the following motion characteristics.

- Which way the mouse (touch pad) moves?
- What is the mouse (touch pad) speed?
- How long is it moved?

Figure 6 shows a mouse (touch pad) movement from the starting point to the end point or click point.

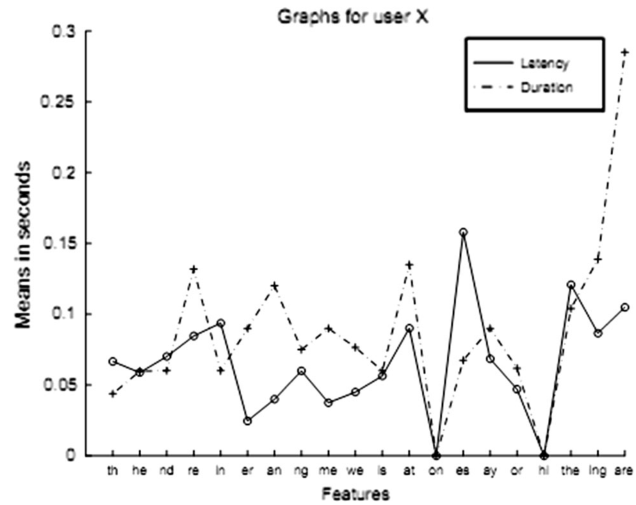

Fig. 5 Sample of recorded keystroke dynamics by latency and duration (A user's fingerprint) [51]

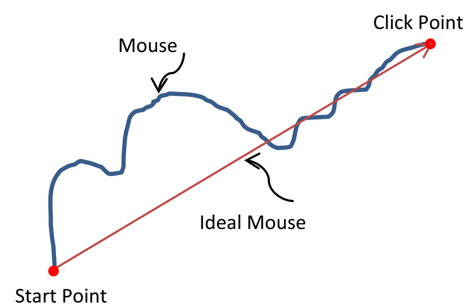

Fig. 6 Example of a movement of mouse (touch pad) without left mouse button pressed. *Red* (straight) vector showing the magnitude and direction of the mouse movement. *Blue* (curved) actual mouse movement (color figure online)

3.2.1 Click: features of the mouse (touch pad) movement

By looking at the mouse (touch pad) movement example in Fig. 6, we can simulate the situation and the total distance travelled via corresponding screen coordinates. This stretch of *click to click* represents the shortest distance between start and end points of the mouse movement on the screen.

3.2.2 Characteristics of the place

This curve extracted from the mouse movement in Fig. 6 is then transformed into a two-dimensional coordinate system; as mentioned previously, the ideal line corresponds to the *x*-axis of the coordinate system, and the *y*-axis therefore describes a measure of the local variation of the mouse (touch pad) movement from the ideal line. In Fig. 7, mouse movement curve, as shown in Fig. 6, has been transformed and is depicted in conjunction with overlay sampling points.

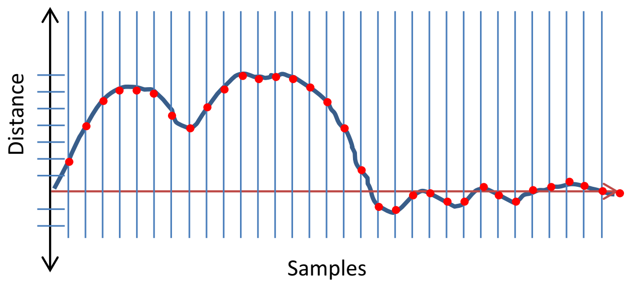

Fig. 7 Real spaces designed for mouse (touch pad) movement ideal line

Then, the transformed curve can be used to evaluate the captured (x, y) coordinate of the mouse movement and the resulting distance values, which are shown in Fig. 7. This newly obtained set of distance values is an expression of the local deviation of the motion from the constructed ideal line.

Since these distance values have lost the absolute commitment to its original screen position (but not its sign, based on the ideal line), it can already measure global properties of the local mouse movement. For example, the sum over all possible distance values states how much the mouse was moving entirely above or below the ideal line. The studied properties are as follows:

- The length of the racing line from start to end point as shown in Fig. 6
- The sum over all distance values
- The zero crossings

The above distances provide individual information about the type of mouse movements giving the following global properties:

- Maximum deviation of the values.
- Average amount of the individual values.
- Standard deviation.
- Variance.

Finally, there are some derived features for mouse movements, which are also considered for more precise recognition of emotions and their calculations:

- Correlation function of the curve.
- First-order and second-order derivatives with their specific evaluation.

It should be noted that the following features can be extracted based on the first and second derivatives of the identified and utilized emotion recognitions:

- Minimum and maximum of the values.
- Average amount over all values.
- Standard deviation.
- Variance.
- Autocorrelation function.

3.2.3 Time properties

In parallel with the above-discussed features, the time intervals, which register with the result of a new (x, y) point, are analyzed. It should not be forgotten that only a change in the x- or y-coordinate of a new data value is read. This elapsed time between two consecutive points not only represents the total time of the mouse movement, but also describes information about the individual movements. Later, the next section explains the time between jerky and slow motions. Also, it can be used to distinguish verse breaks in existing movements very well. A complete overview of all the examined features is firstly presented in Fig. 8. It shows a possible sequence of values of time intervals, from which the main features are very well seen. This figure presents the time between the clicks. This figure is only a demonstration of a sample of registered mouse keystrokes.

As can be seen initially, it is similar to the local variation, made and analyzed with a number of time delta values. Then, first two statements about the time relationships are possible:

- Total time of motion by summing over all values
- Average time distance between two points or the average required time.

However, when a change occurs to the location coordinates of the mouse movements, then averaging is performed on the following:

- Standard deviation of individual values.
- Variance of these values.

Finally, the derived variables of mouse movements and keystrokes are as follows:

- Correlation function.
- First derivative.
- Second derivative with the corresponding analysis.

The formation of a distribution function of these values and the derived properties of this distribution function lead to the catalogue of properties and thus the distribution

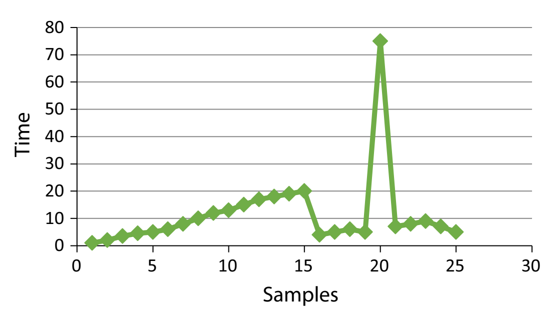

Fig. 8 Elapsed time between the modified coordinates of mouse keystrokes

function. The first derivation of the distribution function with a corresponding evaluation is also added.

3.3 Touch screen (single touch)

The touch screen is only able to determine points in x-, y- and z-coordinates. User interaction with touch screen monitors only results the changes in these coordinate values [48, 63]. These values along with the time interval of changes are collected and prepared to measure the other important features such as velocity and movement details. Recently, there are technologies that utilize user's eye movement, hand gestures and other behaviors to enhance touch capabilities of a touch screen monitor. All of these advanced technologies are the combination of image processing and advanced AI techniques with the touch screen monitors. However, human emotions recognition in touch screen technology is in its initial stages of its development. It is expected that the better techniques will be invented in the future to capture human emotions more dynamically and accurately.

The most significant expansion was therefore to complement the additional available z-component, which has been evaluated in parallel. Thus, analogous reads the (x, y) coordinates of an initial set of z-values, where they open up a value range between 0 and 255. Based on this history, the following features are accessible:

- Average of all z-values.
- Minimum value.
- Maximum value.
- Standard deviation.
- Variance.
- First derivatives.
- Second derivatives.
- Correlation function.

Straight from the emerging contours of the first and second derivatives as well as the correlation function, above additional values can be used to interpret better. These values will be recorded and played to the crowd with the observed features.

By considering this number of features on touch screen monitors, all the values are obtained from the Cartesian coordinate system. However, a three-dimensional coordinate space is presented. This can also offer a transformation in spherical coordinates (r, α, β).

4 Evaluation

This section demonstrates the diagnosis of the research based on the theories and methods of research methodology.

4.1 Evaluation criteria

Evaluation of the system is based on the emotions recognition methods and machine learning techniques, which have been used in the affective system. There are several criteria to evaluate and measure the performance of the system. These criteria are listed in Table 2.

Classification/recognition accuracy is the value that shows how precise a system is able to recognize the emotions. It is mostly focused on the output of machine learning techniques. This criterion is measured by the machine learning classification methods. Generally, for this purpose, from 60 to 80 % of the data would be trained, and then, the rest of 20–40 % of the remaining data would be tested. In the testing section, machine learning method ignores the recorded emotions and predicts/estimates them according to the training section, and later, the final result of recognized emotions (estimated emotions) would be compared to the real recorded emotions. The percentage rate of true classified/recognized emotions is known as recognition accuracy. This value can be represented either as a range of [0, 1] or as percentage. Higher classification accuracy shows higher performance in recognition process.

False-positive rate represents false classified emotions. These emotions are recognized, but they are not matched with the recorded emotions, which are known as false-positive. The resulted percentage of these classified data in the whole dataset is called as false-positive rate. It has the same range of recognition accuracy and can be presented in percentage as well. Despite the recognition accuracy, the lower false-positive ate shows less false detection for emotions, therefore a system having a better performance [64].

Computational time is a classification procedure that takes some time to be applied on the collected dataset. Different classifiers follow different algorithms, and they have different time complexity. The time that is taken for each classification from the beginning to the end is called *computational time* or *process time*. This criterion is mostly related to the nature of the machine learning method and

Table 2 Criteria of evaluation for the recognition methods

No.	Criteria	Description
1	Classification/recognition accuracy	This shows how accurate a system recognizes the emotions
2	False-positive rate	This parameter shows the failure rate in detection
3	Computational time	The required time for the process to be completed

Table 3 SVM: confusion matrix with the normalized values for mouse (touch pad) features classification

Intended emotions	Detected emotions			
	Neutral	Fright	Sadness	Nervousness
Neutral	**0.930**	0.022	0.028	0.020
Fright	0.202	**0.787**	0.004 (0.040)	0.007
Sadness	0.061 (0.084)	0.012	**0.912**	0.015
Nervousness	0.085 (0.175)	0.015	0.065	**0.835**

Bold values are the "true positives" and the rest are "false positives"

the amount of data for training and testing iterations. This parameter is measured in seconds, and lower values show faster emotion recognition [65].

Research studies are trying to improve these values by increasing the recognition accuracy, decreasing false-positive rates and computational time [66].

4.2 Data analysis and evaluation

Evaluations and data analysis were based on the collected data from the prototype software. This section presents different evaluations step by step from classic emotions recognition with the previously discussed features to answer the questions that have been raised in the earlier sections.

The recognition performance is determined by using SVM (with nonlinear Gaussian Kernel) and ANN as classifiers in terms of classification accuracy, false-positive and false-negative rates.

4.2.1 Normalized and maximum classification results

The experiments for every entry in the confusion matrix (Tables 3, 4, 5, 6, 7, 8) are done independently. Therefore, the sum of the recognition accuracy and their relative false-negative values may be more and/or <100 %. To report more reliable results, the false-negative rates have been normalized. The numbers in parenthesis (for SVM classification) are the maximum possible error rates (false-positive and false-negative rates), which have been measured on many iterative runs of the SVM classification [67]. The maximum and normalized values show the possible inaccuracies in the experiment. However, classification experiment on a more homogenous dataset would result less difference among minimum, maximum, average and normalized values of the classification.

4.2.2 Keyboard keystroke dynamics

The number of mistakes in typing (backspace + delete key) was calculated as it shows the proficiency of user in typing. These mistakes can reflect the user's emotions. There are many different methods to correct the mistakes (e.g., selection with the mouse and replacement with keystrokes). It was not possible to catch all of the possible correction scenarios as keystrokes were collected from different applications environments and so we did not have any control on them. However, this feature does give a general idea of the number of mistakes being made.

Outliers for all of the features that involved multiple keys were calculated to remove these pauses (e.g., digraph and trigraph latencies). They were removed by considering the mean and standard deviation for all keystroke dynamic

Table 4 ANN: confusion matrix with the normalized values for mouse (touch pad) features classification

Intended emotions	Detected emotions			
	Neutral	Fright	Sadness	Nervousness
Neutral	**0.911**	0.042	0.015	0.032
Fright	0.333	**0.631**	0.024	0.012
Sadness	0.025	0.102	**0.852**	0.021
Nervousness	0.020	0.097	0.142	**0.741**

Bold values are the "true positives" and the rest are "false positives"

Table 5 SVM: confusion matrix with the normalized values for touch screen

Selected emotion	Detected emotion			
	Neutral	Fright	Sadness	Nervousness
Neutral	**0.710**	0.178 (0.321)	0.090	0.022
Fright	0.015	**0.900**	0.073	0.012
Sadness	0.008	0.099 (0.113)	**0.893**	0.000
Nervousness	0.071	0.354	0.022	**0.553**

Bold values are the "true positives" and the rest are "false positives"

Table 6 ANN: confusion matrix with the normalized values for touch screen

Selected emotion	Detected emotion			
	Neutral	Fright	Sadness	Nervousness
Neutral	**0.538**	0.322	0.094	0.046
Fright	0.068	**0.830**	0.081	0.021
Sadness	0.130	0.081	**0.781**	0.008
Nervousness	0.164	0.203	0.102	**0.531**

Bold values are the "true positives" and the rest are "false positives"

Table 7 SVM: confusion matrix with the normalized values of keyboard, mouse (touch pad) and touch screen

Selected emotion	Detected emotion			
	Neutral	Fright	Sadness	Nervousness
Neutral	**0.851**	0.121	0.006 (0.076)	0.022
Fright	0.001	**0.932**	0.055 (0.082)	0.012
Sadness	0.009	0.064 (0.118)	**0.921**	0.006
Nervousness	0.092	0.261	0.087 (0.122)	**0.650**

Bold values are the "true positives" and the rest are "false positives"

Table 8 ANN: confusion matrix with the normalized values of keyboard, mouse (touch pad) and touch screen

Selected emotion	Detected emotion			
	Neutral	Fright	Sadness	Nervousness
Neutral	**0.883**	0.018	0.053	0.046
Fright	0.011	**0.810**	0.137	0.042
Sadness	0.071	0.039	**0.807**	0.083
Nervousness	0.075	0.071	0.092	**0.762**

Bold values are the "true positives" and the rest are "false positives"

features, which they were 12 standard deviations greater than the mean for each individual participant [53]. This process has been considered in the prototype application while recording and collecting the data from users.

The percentage values in Fig. 9 show the classification results of human emotions based on the keyboard keystroke dynamics with their false-positive rates by ANN and SVM. Fright emotion has the strongest classification accuracy of 91.24 % by SVM and the least value of 4.35 % for false-positive rate in SVM classification.

4.2.3 *Mouse (touch pad: single touch)*

Figure 10 shows two primary and simple emotion recognition results separately. In the first detection process, features were selected randomly and the result is shown in a lighter (orange) color. Clearly, the best result was on Neutral emotion, but for the other emotions, the outcome is <40 %. Then, in the second round, the features were selected according to the Schuller [48], and it turned to the blue results that are far better than the first result. In contrast to the weak result percentage of *fright* (*afraid*) emotion, the other two emotions of *pensive* and *annoyed* results are remarkably improved. There are two exemplary emotions that are clearly satisfied in this network diagram, which are *pensive* and *annoyed*. Depending on the values at zero, the more dissatisfied section was the subject of recognition of the corresponding emotion. But neutral was a very different emotion, apart from the emotion category for the time being. Each person has their own movement and behavior of the mouse that is a general synthetic dataset which was not individually tailored.

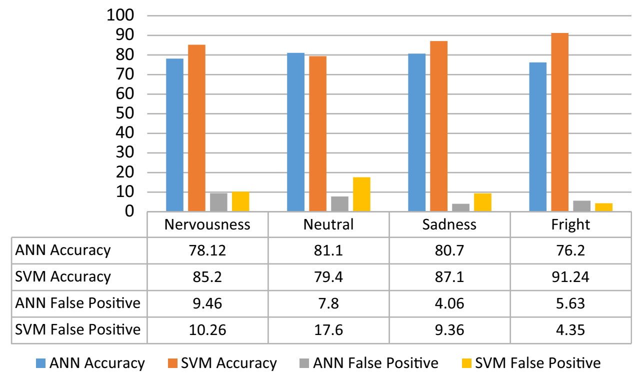

Fig. 9 Keyboard keystroke dynamics classification accuracy and false-positive rates of ANN and SVM

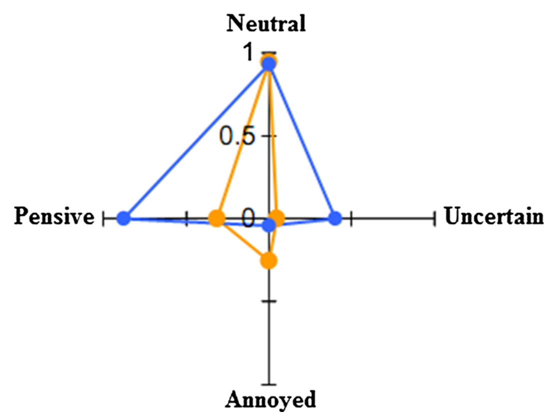

Fig. 10 Percentage of emotions recognition in two phases by changing the features

In details, the collected data from our volunteers were evaluated; then, it is tried to analyze with their emotions. Tables 3 and 4 and Figs. 11 and 12 summarize the evaluation of all collected datasets on four sets of emotions. This evaluation was performed based on the collected data vectors [68, 69] and written in a confusion matrix. Each emotion set was trained and evaluated separately. The overall average of the correctly classified emotions is 0.866 with a mean variance of 0.075. In summary, the following two precision characteristics in determining the results are particularly striking:

- The highest precision is recognized in the neutral and annoyed categories.
- The lowest precision is for the *fright (afraid)* category.

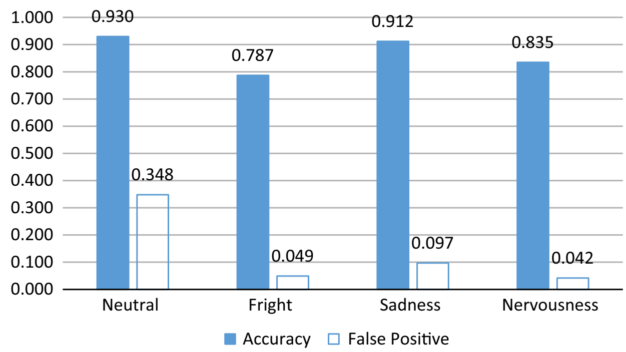

Fig. 11 SVM: mouse (touch pad) features classification accuracy and false-positive rates

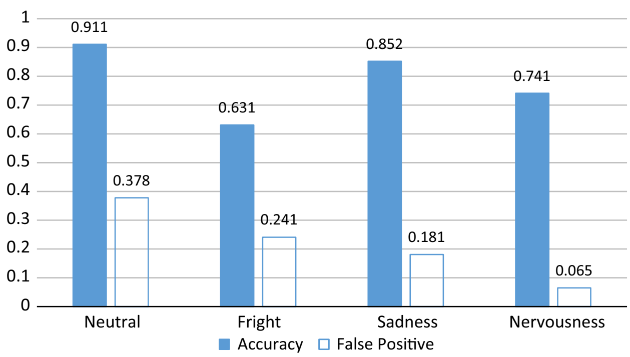

Fig. 12 ANN: mouse (touch pad) features classification accuracy and false-positive rates

The correctly classified emotions by RapidMiner are the values at the junction of the same detected emotion with the intended emotion as shown in bold in Table 3 and Table 4. The other values are false-positive alarms, which are classified incorrectly. The increased classification rate for neutral emotion can be explained easily.

The test subjects were accumulated primarily with expectations of neutral data vectors for the emotion, because this is probably the most common emotion, which is felt to have been distributed over the days. For example, the collected data by the volunteers at the individual days can be represented neutral with up to 93 % on certain days.

Annoyed caused a different system behavior. If a PC user has the emotion of *annoyed*, he moves the mouse usually very fast and also fixed with short presses on the mouse button. The properties of mouse movement are *fast* and *brief*. In the short press, the mouse pointer does not move. Thus, generally, no movement is detected during the mouse click. Above all, the movement between pressing the mouse button (clicking) gives information about the emotion. The movements are almost in all cases, strongly oriented toward a goal, and thus, the coordinates of the determined distances to the ideal line are rather low.

But a question still remains on why the precision of the detection is still low. The probable answer was found when the analysis on the volunteers was performed. At the time of working in different situations with the computer, they are not sure about their own emotions. When they are asked to identify their emotions, they are rather unsure what kind of emotion they have at the moment.

The most inaccurate result was obtained in *fright (afraid)* emotion. The reason is that just moving the mouse while pressing the mouse button is relevant to this emotion. Thus, an insecure person presses a little longer and deliberates on the mouse, where they will lead, and the person did not intend slight movement of the cursor. Data analysis of the features is shown for the recognition of emotion, and it is not very meaningful, and this is probably one of the reasons for the lower values in the confusion matrix.

Finally, it can be concluded that although the recognition of emotion in sufficient degree happens, unfortunately the lack of standard hardware with significant qualities causes a lower accuracy. It would be very important to have several data collection periods to increase the strength of the data. It also brings more clarity about the emotions, and it enables better detection.

4.2.4 *Touch screen (single touch)*

In the signal processing field, it is already known that the auto correlation function (ACF) is the function of $\varphi(T)$ by a measure of the *inner context* of a signal $s(t)$. Therefore, it is a measure of similarity or correlation of signal sections. In the other words, τ is a shifted time against each other [70]:

$$\varphi(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)s(t+\tau)\mathrm{d}t \tag{1}$$

The ACF is recorded directly as the temporal correlation function of a time series. There is a context in each one, and it is examined the same measure at different time points and plotted as a function of the difference. This feature implies as the *memory* of the system. The ACF consists of only a single peak at $\tau = 0$ and disappears for all other τ values. Then, the series of measurements can be analyzed in a completely reliable, stochastic behavior.

Thus, for $\tau = 0$, the internal consistency of the signal is the greatest, and the resulting value is an addition with the power of the signal $s(t)$. For $\tau > 0$, the value of the ACF is small, with $\varphi(\tau)$ while $\tau \to \infty$ tends to zero when $s(t)$ is neither a constant nor contains a periodic component, especially when voice signals are more pronounced by negative correlation values of significance. For example, with a male speaker for ms5 $\approx \tau$ significant negative values, this is at a fundamental frequency of the speaker of 100 Hz due to the shift by a half wave.

Finally, ACF can suppress the signals, which are superimposed by a noise, and their noise signal pulse duration is very small compared with the signals to interference suppression. From the theory of the noise, it is apparent that particular white noise can be suppressed by the autocorrelation method. This is possible even if the measuring signal in time domain is no longer detected in the noise because it goes down.

4.2.5 ACF on z-values

Now, it would be an attempt to transfer the meaning of the ACF in a speech signal to the resulting data series. The broadest sense is a kind of power function with the corresponding z-values. A very slow pressure on the screen surface and also harmonic pressure reduction lead to a significantly higher value of the ACF at $\tau = 0$, such as a very jerky, short pressure pulse. Similarly, the mentioned strong negative values can achieve a magnitude of much higher value than the short and strong pulses by a slow press on the screen.

It also seems plausible that PC users tend to be at a certain emotion when they touch the screen in the same way, but each time you touch it too shortly may result the system recognition inaccuracies. This fact can now be compared with the noise as previously mentioned. The small deviations overlap is registered as a white noise with a kind of *pure basic series of z-values*, which is typical for a specific emotion. It can be processed by using the ACF with a particular degree of strong noisy signals. This is a striking indication of the importance of the ACF for the components of a data series.

Tables 5, 6 and Figs. 13, 14 show the final normalized results over all the test subjects. The overall confusion matrix with an overall mean value of 0.76 (76 %) for the correctly classified emotions, as shown in bold font, is achieved. After the evaluation of the existing system for the detection of the four emotions, with their detailed explanations, it can be concluded that this system can be used for emotion recognition with an acceptable accuracy.

4.3 Hybrid of keyboard–mouse (touch pad)–touch screen

After each individual evaluation on analysis of keyboard keystroke dynamics, mouse movements and touch screen interaction, a hybrid model is tested. In the hybrid model, the accuracy of the *fright (afraid)* emotion is the best among the others. Neutral and Nervousness have the lowest accuracies, and these two emotions have the greatest rate of confusion with each other. These results are tabulated in Table 7.

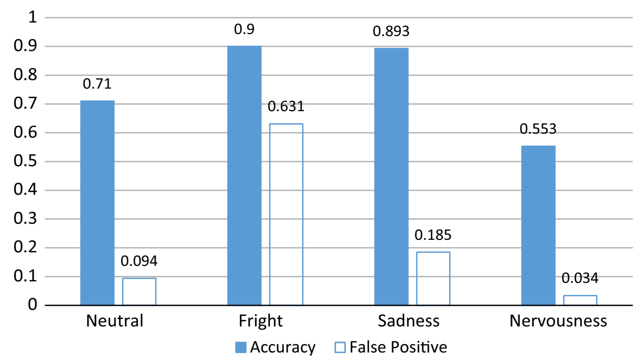

Fig. 13 SVM: touch screen features classification accuracy and false-positive rates

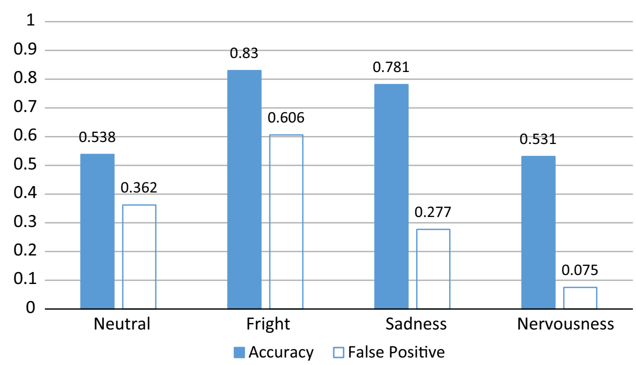

Fig. 14 ANN: touch screen features classification accuracy and false-positive rates

As it can be seen in Tables 7, 8 and Figs. 15, 16, all of the four emotions have been detected more accurately by using a hybrid and combination of all three input devices (keyboard, mouse and touch screen). Also in some cases, the error has been escalated a little bit, but the increase in performance is much higher than the error rates (false-positive and false-negative rates).

5 Discussion

The three evaluation criteria used in this study were as follows:

- The recognition accuracy is the most important criterion in the evaluation. The proposed methods of this research have been evaluated in terms of classification/recognition accuracy; then, at the end, they are compared with the similar research areas in measuring human emotions recognition accuracy in computing.
- The false-positive rate has been shown in every confusion table. However, the lack of enough

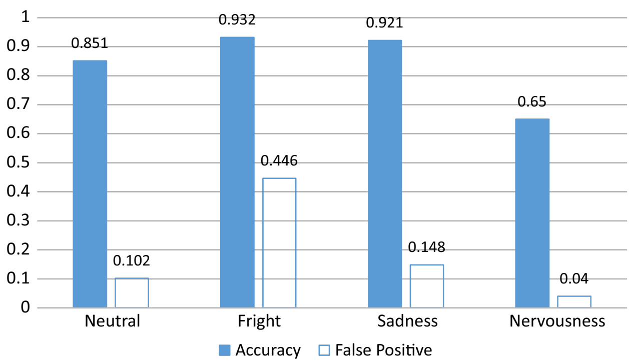

Fig. 15 SVM: keyboard, mouse and touch screen features classification accuracy and false-positive rates

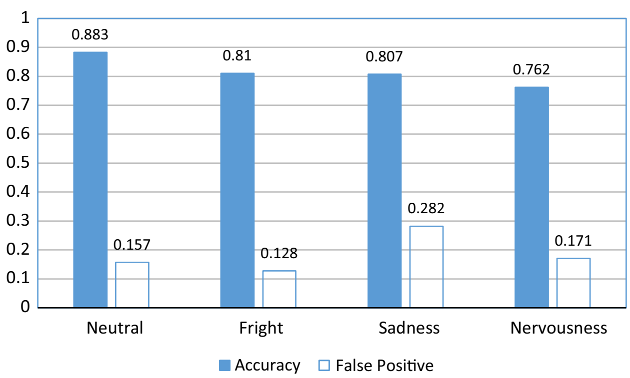

Fig. 16 ANN: keyboard, mouse and touch screen features classification accuracy and false-positive rates

information in the previous research papers, comparing the results of this study with the similar works, was not possible.

- The computational/processing time. This is only related to the classification methods and the number of extracted features.

Previously researchers tried to gain more precise results in affective computing, while some fails to consider all different aspects of efficiency, usability and real-time performance. This research tried to obtain a reliable method which considers all these aspects and is more suitable to be applied in real-world applications. It could gain a high accuracy in comparison with the other scholarly research, especially by combining three methods together. A reliable high accurate result of 93.20 % by SVM classification method has been achieved, which is competitive with all other previous methods. However, facial expressions stands on the top, yet our hybrid method could fulfill deficiency and limitations of image processing techniques. Figure 17 compares the achieved result with the other methods and accuracies discussed earlier.

The final result of 93.20 % was compared with the superior research on affective computing with different

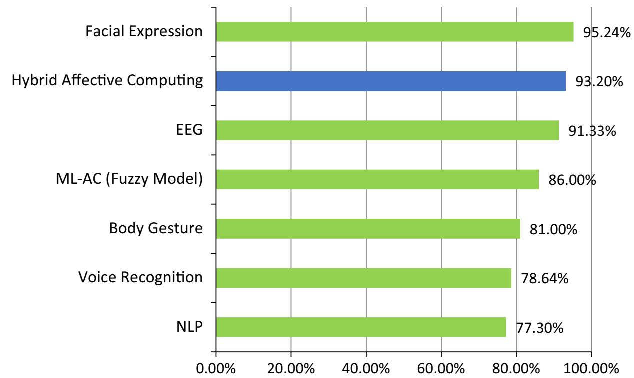

Fig. 17 Comparison of the best accuracies of different methods for affective computing

methods. An affective system based on EEG has gained the accuracy of 66 % in the years of 2009 and 2010 [44, 45]. In 2013, despite other similar research by EEG, 91.33 % of accuracy is achieved by using KNN machine learning method [47]. Regarding NLP, in 2008, Lei et al. [31] got the result of 74.1 % in the emotion of *anger*; Calvo and Kim [33] achieved 77.30 % in anger/disgust and Joy by using CNMF method. Voice processing in emotion recognition achieved the 78.64 % of accuracy [39, 71]. Affective systems based on body gesture recognition have been resulted in 81 % by Gunes and Piccardi [27]. The most similar approaches to this research method have been done by Milanova and Sirakov [49]; they gained 87 % of emotions recognition accuracy. The best competitive method is facial expression recognition, which has been improved a lot. Among many researchers in facial expression recognition, Konar et al. [16], Kao and Fahn [17] and Beaudrya et al. [20] got 88.68, 90 and 95.24 %, respectively, of accuracy. Figure 17 shows that the methods employed in this research have resulted roughly 6 % better than the similar methods.

6 Conclusion and future work

This research has proposed a hybrid solution for human emotion recognition (affective computing), which is outstanding from different aspects. Firstly, the recognition accuracy by using SVM classification shows a reliable 93.20 % of accuracy at the best of its recognition rate in this study. Secondly, the computational time of this methodology is much less than the other techniques such as image and signal processing. Thirdly, the required hardware in this research is mostly available in all personal computers. Finally, this research has eliminated the discussed privacy issues by using camera and microphone.

To name a few limitations of this research, we can refer to the new input devices such as multi-touch touch pads and multi-touch touch screen displays, which are not covered in this scope. Human emotion is a cognitive concept and is hard to recognize the exact emotion with its degree. This research collected the data based on the users' Self-Reported emotion. It is possible that the users could not identify their own emotions properly, which may alter the results. Furthermore, the achieved result is limited to some few groups of emotions, which are common in the similar research to make a comparative analysis. However, other emotions were not investigated. At last, evaluation was conducted based on only three criteria, which can be extended in future to cover more aspects of efficiency and usability.

Many researchers in affective computing and psychology proved that there are some minor differences in the definitions and expressions of each emotion among people all around the world with different language and cultural backgrounds [7]. Participants in the research showed only the emotional features on themselves, and the resulted data may not be extended to the other people from other part of the world. Lack of a standard framework on emotions for people in different regions is one of the critiques of this research and the whole field of research for affective computing. Lewis has discussed the strength of our hypothesis to cluster the collected data based on the language and cultural backgrounds to differentiate the results based on the emotions [7].

To address the above issues, we are currently working on a fuzzy model for affective computing to detect other available emotions (not only one emotion) with their dominance values [72, 73]. This fuzzy model tries to cluster and classify the emotions based on the cultural differences. Building a universal model to be extended and used on all users globally is an open challenge in affective computing [74, 75].

We also plan to implement this method on new devices with new features and platforms such as smart phones. Result of this research is going to be used in a recommender system [74] in a way that system adjusts its recommendation according to user's emotions and interactions.

Acknowledgments Thanks to the students of Universiti Kebangsaan Malaysia (The National University of Malaysia), University of Duisburg-Essen, Germany, and members of AIESEC UKM and AIESEC ESSEN, the largest nonprofit organization run by students, who participated in the prototype system evaluation.

References

1. Leon E, Clarke G, Callaghan V, Sepulveda F (2007) A user-independent real-time emotion recognition system for software agents in domestic environments. Eng Appl Artif Intell 20(3):337–345. doi:10.1016/j.engappai.2006.06.001
2. Harris M, Ishii H, Chung C, Dodsworth C, Buxton B (1999) Natural and invisible human interfaces. In: International conference on computer graphics and interactive techniques (SIGGRAPH). doi:10.1145/311625.311922
3. Palm G, Glodek M (2013) Towards emotion recognition in human computer interaction. In: Neural nets and surroundings, vol 19. Springer, pp 323–336. doi:10.1007/978-3-642-35467-0_32
4. Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG (2001) Emotion recognition in human–computer interaction. Sig Process Mag IEEE 18(1):32–80
5. Picard RW (1998) Affective computing. MIT Press, Cambridge
6. Strongman KT (2003) The psychology of emotion. Department of Psychology, University of Canterbury, 5th edn. Wiley, Christchurch
7. Lewis M, Haviland-Jones JM, Barrett LF (2010) The cultural psychology of the emotions—ancient and renewed. In: Richard A, Shweder JH, Horton R, Joseph C (eds) Handbook of emotions, 3rd edn. The Guilford Press, New York, p 19
8. Fredrickson BL (2001) The role of positive emotions in positive psychology—the broaden-and-build theory of positive emotions. Am Psychol 56(3):9. doi:10.1037/0003-066X.56.3.218
9. SRamaraj V, Ravindran A, Thirumurugan A (2013) Emotion recognition from human eye expression. IJRCCT 2(4):158–164
10. Schmidt K, Cohn J (2002) Human facial expressions as adaptations: evolutionary questions in facial expression. Am J Phys Anthropol 33:3–24
11. Baker C, Cokely D (1980) American sign language: a teacher's resource text on grammar and culture. T.J. Publishers, Silver Spring
12. Jamshidnejad A (2009) Facial emotion recognition for human computer interaction using a fuzzy model in the e-business. In: Conference on innovative technologies in intelligent systems and industrial applications (CITISIA), pp 202–204
13. Tsihrintzis GA, Virvou M, Alepis E, Stathopoulou IO (2008) Towards improving visual-facial emotion recognition through use of complementary keyboard-stroke pattern information. In: Fifth international conference on information technology: new generations, pp 32–37
14. Valstar M, Patras I, Pantic M (2004) Facial action unit recognition using temporal templates. In: IEEE international workshop on robot and human interactive communication, pp 253–258
15. Gu W, Xiang C, Venkatesh Y, Huang D, Lin H (2012) Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. Pattern Recogn 45(1):80–91
16. Konar A, Chakraborty A, Halder A, Mandal R, Janarthanan R (2012) Interval type-2 fuzzy model for emotion recognition from facial expression. In: Kundu MK, Mitra S, Mazumdar D, Pal SK (eds) Perception and machine intelligence. Lecture notes in computer science, vol 7143. Springer, Heidelberg, pp 114–121. doi:10.1007/978-3-642-27387-2_15
17. Kao CY, Fahn CS (2012) A design of face detection and facial expression recognition techniques based on boosting schema. Appl Mech Mater 121:617–621
18. Ilbeygi M, Shah-Hosseini H (2012) A novel fuzzy facial expression recognition system based on facial feature extraction from color face images. Eng Appl Artif Intell 25(1):130–146. doi:10.1016/j.engappai.2011.07.004
19. Rahulamathavan Y, Phan RC-W, Chambers JA, Parish DJ (2013) Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. IEEE Trans Affect Comput 4(1):83–92. doi:10.1109/T-AFFC.2012.33
20. Beaudrya O, Roy-Charlanda A, Perrona M, Cormierb I, Tappa R (2014) Featural processing in recognition of emotional facial expressions. Cogn Emot 28(3):416–432. doi:10.1080/02699931.2013.833500

21. Kleinsmith A, Bianchi-Berthouze N (2013) Affective body expression perception and recognition: a survey. IEEE Trans Affect Comput 4(1):15–33. doi:10.1109/T-AFFC.2012.16
22. Lee JS, Shin D-H (2013) A study on the interaction between human and smart devices based on emotion recognition. In: HCI international 2013-posters' extended abstracts. Springer, pp 352–356
23. Bulwer J (1644) Chirologia: or the natural language of the hand. London. doi:10.1037/11828-001
24. de Jorio A (1832) Gesture in naples and gesture in classical antiquity. Indiana University Press, Bloomington
25. McNeill D (2005) Gesture and thought. Chicago University Press, Chicago
26. Goldin-Meadow S (2003) Hearing gesture: how our hands help us think. Harvard University Press, Cambridge
27. Gunes H, Piccardi M (2005) Fusing face and body gesture for machine recognition of emotions. In: IEEE international workshop on robots and human interactive communication, pp 306–311
28. Glowinski D, Camurri A, Volpe G, Dael N, Scherer K (2008) Technique for automatic emotion recognition by body gesture analysis. In: IEEE computer society conference on computer vision and pattern recognition workshops
29. Chena S, Tiana Y, Liub Q, Metaxasc DN (2013) Recognizing expressions from face and body gesture by temporal normalized motion and appearance features. Image Vis Comput 31(2):175–185. doi:10.1016/j.imavis.2012.06.014
30. Steunebrink BR, Dastani M, Meyer JJC (2009) The OCC model revisited. Utrecht University, The Netherlands
31. Li H, Pang N, Guo S, Wang H (2008) Research on textual emotion recognition incorporating personality factor. In: IEEE international conference on robotics and biomimetics, pp 2222–2227
32. Gil GB, Jesús ABd, Lopéz JMM (2013) combining machine learning techniques and natural language processing to infer emotions using Spanish twitter corpus. In: International workshops of PAAMS 365:149–167. doi:10.1007/978-3-642-38061-7_15
33. Calvo RA, Kim SM (2013) Emotions in text: dimensional and categorical models. Comput Intell 29(3):527–543. doi:10.1111/j.1467-8640.2012.00456.x
34. Shahin I (2013) Speaker identification in emotional talking environments based on CSPHMM2s. Eng Appl Artif Intell 26(7):1652–1659. doi:10.1016/j.engappai.2013.03.013
35. Klecková J (2009) Important nonverbal attributes for spontaneous speech recognition. In: Fourth international conference on systems, pp 13–16
36. Wang Y, Guan L (2008) Recognizing human emotional state from audiovisual signals*. Multimed IEEE Trans 10(5):936–946
37. Merrett T (2008) Query by humming. McGill University, Montreal
38. Unal E, Chew E, Georgiou PG, Narayanan SS (2008) Challenging uncertainty in query by humming systems: a fingerprinting approach. Audio Speech Lang Process IEEE Trans 16(2):359–371
39. Amarakeerthi S, Ranaweera R, Cohen M (2010) Speech-based emotion characterization using postures and gestures in CVEs. In: International conference on cyberworlds, pp 72–76
40. Pittermann J, Schmitt A, El Sayed NF (2008) Integrating linguistic cues into speech-based emotion recognition. In: 4th international conference on intelligent environments, pp 1–4
41. Huang T (2008) Audio-visual human computer interface. In: IEEE international symposium on consumer electronics, pp 1–1
42. Kotti M, Kotropoulos C (2008) Gender classification in two emotional speech databases. In: 19th international conference on pattern recognition, pp 1–4
43. Krishnan A, Fernandez M (2014) System and method for recognizing emotional state from a speech signal. Google Patents
44. Liu Y, Sourina O, Nguyen MK (2010) Real-time EEG-based human emotion recognition and visualization. In: International conference on cyberworlds, pp 262–269
45. Schaaff K, Schultz T (2009) Towards emotion recognition from electroencephalographic signals. In: Affective computing and intelligent interaction and workshops (ACII), pp 1–6
46. Guangying Y, Shanxiao Y (2010) Emotion recognition of electromyography based on support vector machine. In: Third international symposium on intelligent information technology and security informatics, pp 298–301
47. Murugappan M, Murugappan S (2013) Human emotion recognition through short time electroencephalogram (EEG) signals using fast fourier transform (FFT). In: IEEE 9th international colloquium on signal processing and its applications (CSPA), pp 289–294. doi:10.1109/CSPA.2013.6530058
48. Schuller B, Rigoll G, Lang M (2004) Emotion recognition in the manual interaction with graphical user interfaces. In: IEEE international conference on multimedia and expo 2. pp 1215–1218
49. Milanova M, Sirakov N (2008) Recognition of emotional states in natural human–computer interaction. In: IEEE international symposium on signal processing and information technology (ISSPIT), pp 186–191
50. Kaklauskas A, Zavadskas EK, Seniut M, Dzemyda G, Stankevic V, Simkevičius C, Stankevic T, Paliskiene R, Matuliauskaite A, Kildiene S, Bartkiene L, Ivanikovas S, Gribniak V (2011) Web-based biometric computer mouse advisory system to analyze a user's emotions and work productivity. Eng Appl Artif Intell 24(6):928–945. doi:10.1016/j.engappai.2011.04.006
51. Monrose F, Rubin AD (2000) Keystroke dynamics as a biometric for authentication. Future Gener Comput Syst 16(4):351–359
52. Wang R, Fang B (2008) Affective computing and biometrics based HCI surveillance system. In: International symposium on information science and engineering, pp 192–195
53. Epp C, Lippold M, Mandryk RL (2011) Identifying emotional states using keystroke dynamics. In: Proceedings of the 2011 annual conference on human factors in computing systems, pp 715–724. doi:10.1145/1978942.1979046
54. Waikato Uo (2010) Weka. http://www.cs.waikato.ac.nz/ml/weka/. Accessed 01 July 2010
55. Bashir MG, Nagarajan R, Hazry D (2010) Facial emotion detection using GPSO and Lucas–Kanade algorithms. In: International conference on computer and communication engineering (ICCCE), pp 1–6. doi:10.1109/ICCCE.2010.5556754
56. Ekman P, Friesen WV (2003) Unmasking the face: a guide to recognizing emotions from facial clues. Malor Books, Cambridge
57. Kao ECC, Liu CC, Yang TH, Hsieh CT, Soo VW (2009) Towards text-based emotion detection. In: International conference on information management and engineering (ICIME), pp 70–74
58. Yang H, Willis A, De Roeck A, Nuseibeh B (2012) A hybrid model for automatic emotion recognition in suicide notes. Biomed Inform Insights 5(Supp 1):17–30. doi:10.4137/BII.S8948
59. Böhlen M (2009) Second order ambient intelligence. J Ambient Intell Smart Environ 1(1):63–67
60. Aarts E, de Ruyter B (2009) New research perspectives on ambient intelligence. J Ambient Intell Smart Environ 1(1):5–14
61. Sinek S (2011) First why and then trust. TEDx. http://sciencestage.com/v/42756/tedxmaastricht-simon-sinek-first-why-and-then-trust.html. Accessed 06 Jan 2012
62. Larson R, Csikszentmihalyi M (1983) The experience sampling method. In: New directions for methodology of social & behavioral science, vol 15, pp 41–46

63. Hertenstein MJ, Keltner D, App B, Bulleit BA, Jaskolka AR (2006) Touch communicates distinct emotions. Emotion 6(3):528
64. Rezaee Jordehi A (2014) A chaotic-based big bang–big crunch algorithm for solving global optimisation problems. Neural Comput Appl 25(6):1329–1335. doi:10.1007/s00521-014-1613-1
65. Rezaee Jordehi A (2014) Particle swarm optimisation for dynamic optimisation problems: a review. Neural Comput Appl 25(7–8):1507–1516. doi:10.1007/s00521-014-1661-6
66. Han J, Kamber M, Pei J (2012) Data mining concepts and techniques, 3rd edn. Elsevier, USA
67. Bakhtiyari K, Taghavi M, Husain H (2014) Implementation of emotional-aware computer systems using typical input devices. In: Nguyen NT, Attachoo B, Trawiński B, Somboonviwat K (eds) Intelligent Information and Database Systems, vol 8397. Springer International Publishing, Bangkok, pp 364–374. doi:10.1007/978-3-319-05476-6_37
68. Kemp F (2003) Applied multiple regression/correlation analysis for the behavioral sciences. J R Stat Soc Ser D (Stat) 52(4):691
69. Cohen J (2003) Applied multiple regression/correlation analysis for the behavioral sciences. Lawrence Erlbaum Associates, Publishers
70. Hauske G (2003) Statistische Signaltheorie. In: Skriptum zur Vorlesung. Technische Universität München, München, Deutschland
71. Xiao Z, Dellandrea E, Dou W, Chen L (2007) Automatic hierarchical classification of emotional speech. In: Ninth IEEE international symposium on multimedia workshops, pp 291–296
72. Bakhtiyari K, Husain H (2013) Fuzzy model on human emotions recognition. In: 12th international conference on applications of computer engineering, pp 77–82
73. Bakhtiyari K, Husain H (2014) Fuzzy model of dominance emotions in affective computing. Neural Comput Appl 25(6):1467–1477. doi:10.1007/s00521-014-1637-6
74. Taghavi M, Bakhtiyari K, Scavino E (2013) Agent-based computational investing recommender system. In: Proceedings of the 7th ACM conference on recommender systems, pp 455–458. doi:10.1145/2507157.2508072
75. Taghavi M, Bakhtiyari K, Taghavi H, Olyaee V, Hussain A (2014) Planning for sustainable development in the emerging information societies. J Sci Technol Policy Manag 5(3):178–211. doi:10.1108/JSTPM-04-2014-0013