Investigating Keystroke Dynamics and their Relevance for Real-time Emotion Recognition

Aicha Maalej[a,*], Ilhem Kallel[a,c], Javier Jesus Sanchez Medina[b]

[a]*REGIM-Lab: Research Groups in Intelligent Machines, ENIS University of Sfax, Tunisia*
[b]*CICEI: Innovation Center for the Information Society, University of Las Palmas de Gran Canaria, Spain*
[c]*ISIMS: Higher Institute of Computer Science and Multimedia, University of Sfax, Tunisia*

Abstract

There is strong evidence that emotional states affect the Human's performance and decision making. Therefore, understanding Human emotions has become of great concern in the field of Human Computer Interaction (HCI). One way to online emotion recognition is through Keystroke Dynamics. It addresses the drawbacks of current methods which are intrusive and not user-friendly, expensive to implement, and neither realistic nor applicable in a real-world context. The keystroke dynamics approach focuses on analyzing the particular way a person types on a keyboard. In our research work, we start by developing a web application (EmoSurv) in order to collect the data and build a dataset. We generate datasets for free-text and fixed-text entries. These datasets are labeled with emotional states of the participants (Angry, Happy, Sad, Calm, and Neutral state). The obtained datasets are used for training and building models using machine learning algorithms. Outstanding accuracy rates are obtained reaching 93.922 % and Kappa equal to 0.9197 using Random Committee algorithm. We finally provide a set of recommendations for future experimentation by comparing the different models generated.

Keywords: Keystroke Dynamics, Affective Computing, Human Computer Interaction, Emotion Recognition, Data Collection, Machine Learning

*aysha.maalej@gmail.com

1. Introduction

People can sense each other's emotions while communicating and interacting. Understanding one's current emotion helps in managing situations, in leading more reliable communications and in decision making. In view of the tremendous emergence of digitalization and the advancement of computer systems, people's interactions with digital devices should be more investigated. According to Picard [1] , if we want computers to be genuinely intelligent and to interact naturally with us, we must give them the ability to recognize, understand, and even to have and express emotions.

Some of the major functional roles of emotions are: triggering motivated behaviors, enriching communications, and affecting cognitive processing [2]. However, nowadays people spend a great deal of time interacting with their digital devices that overlook their emotions. Computer systems that do not understand or adapt to a user's context, such as his emotions, could fall into the problem of usability. Ignoring a user's emotional state or failing to manifest the appropriate emotion, can drastically hinder performance and risks to be considered as cold, socially inept, untrustworthy, and incompetent [3]. In such situations, emotionally intelligent systems would provide a richer context and help in making intelligent decisions.

In an effort to make computer systems sense emotions, research efforts in affective computing [1] have significantly increased in recent years and researchers have come with different methods like measuring physiological signals [4] or through recording Electroencephalogram (EEG) [5]. Although most of these solutions have provided significant accuracy rates, they are limited in a number of ways. In fact, they demand additional expensive equipment and hardware setup. Besides, in the case of detecting emotions based on physiological states for example [6], equipment used may have intrusive nature [7]; Sensors take time to attach and shaving may be necessary. All these approaches require the use of additional devices that, other than being very expensive to implement, can affect the natural emotional state of a user. Above all this, these methods are not realistic and applicable in a real-world context.

In an attempt to mitigate the limitations of emotion recognition methods, we opt for studying the typing rhythms of computer users to infer their current emotional states. This method is known as Keystroke dynamics. Keystroke dynamics contains sufficient information and it was first exploited in authentication systems [8] where it serves as a potential biometric identifier to ascertain a specific keyboard user.

The main motivation behind using keystroke dynamics is its non-intrusiveness, user-friendliness, low implementation and deployment cost, ease of integration into the computer with minimal user intervention, prompt recognition (real-time), and adaptability in remote devices.

Our purpose in this work is to train and build models based on a dataset that contains keystroke features and emotion labels. For that, we will firstly explore; what features can be extracted from a users particular typing rhythm?, what classification techniques are used in emotion recognition through keystroke dynamics?, and how data is collected and are there any datasets publicly available?

What are emotions?

In psychology, emotions are extensively studied. They exert an incredibly powerful force on human behavior and motivation. Although they are widely used and studied, we cannot find an exact definition of the term "emotion".

Plutchik [9] estimated that more than 90 definitions of emotions were suggested in the 20th century. We report here one of his definitions: *"An emotion is not simply a feeling state. Emotion is a complex chain of loosely connected events that begins with a stimulus and includes feelings, psychological changes, impulses to action and specific, goal-directed behavior. That is to say, feelings do not happen in isolation. They are responses to significant situations in an individual's life, and often they motivate actions."*

Emotions, as explained by Scherer [10], are triggered in an individual by some events that can be either internal or external . He indicates that external events may include the behavior of other people, change of situation or a novel situation, while internal events include thoughts, memories, or sensations.

Keystroke Dynamics

Keystroke dynamics (also known as keystroke biometrics) is a behavioral biometric which describes the typing rhythm of a user on digital devices. These devices may refer to a: computer keyboard, mobile phone, or touch screen panel. In our research, we refer to keystroke dynamics as the process of measuring and assessing human's typing rhythm on a physical keyboard in order to investigate the correlation between emotions and typing patterns.

Keystroke dynamics can also be defined as the exact timing information of each key pressed (Key Down) and key released (Key Up) while a person is typing on his/her keyboard. The main purpose behind this biometric measurement is that typing rhythm is different for every emotional state, like any other biometric, it allows to identify the current emotion of a user.

In the remainder of this paper, we summarize the related work. Then we describe the methodology that we adopt for the data collection process. We also shape and preprocess the dataset. Subsequently, we display and discuss the results obtained after applying the classification algorithms. Finally, a conclusion is drawn summarizing and assessing the contribution of this paper.

2. Literature Review

Emotion recognition is extensively studied in literature. The aim of this work is to study the extend to which emotions can be recognized using keystroke dynamics. To this end, we take a glance at the different techniques used for emotion recognition before deeply investigating how previous studies approached the methodological steps to build an emotion recognition system.

2.1. Keystroke Dynamics over other Techniques

Emotions can be sensed from: facial expressions [11, 12, 13, 14, 15], body gestures and movements [16], physiological states [4, 5, 6, 17, 18, 19, 20, 21], speech and voices [11, 22, 23], text [24, 25, 22], and keyboard and mouse dynamics [26] . Such techniques reveal significant accuracy; however, they demand additional expensive equipment and hardware setup. As for detecting emotions based on physiological states, equipment used may have intrusive nature. All these approaches require the use of additional devices that, other than being not realistic in a casual usage context, they can affect the natural emotional state of a user. In textual emotion recognition, obstacles like spelling errors and slang can be faced. Besides, additional steps must be performed and an ontology, as well as a semantic analysis, is required.

Although Keystroke dynamics is not universal and is not the best in terms of performance, it is obviously the most accepted by users since data is easily collected and such process is not intrusive to the user. For instance, the user does not need to put sensors on his/her body or be exposed to any other machine. Other advantages of using keystroke dynamics are the low implementation and deployment cost, ease of integration into the computer with minimal user intervention, prompt recognition process (real-time), and adaptability in remote devices. Unlike many other biometrics, keystroke data can be collected using only one's personal computer with no additional hardware. However, some weaknesses can be faced; e.g., big amounts of data should be gathered for the training process, and the instability along time; meaning that the values of keystroke parameters taken from a user may depend on the type of the hardware (keyboard) used. Moreover, a possible decrease in accuracy due to the variations in typing rhythm, caused by external factors like injury, fatigue, or distraction, may occur.

2.2. Keystroke Dynamics Applications

Keystroke dynamics seems to provide useful information about the user, thus exploited in many areas other than emotion recognition such as user authentication [27], age estimation [28], gender prediction [29], and authentication [30]. However, it has been proven that keystroke dynamics can exhibit instabilities due to transient factors such as emotion, stress, and drowsiness [31]. These results have motivated researchers, as well as this present research, to investigate more the effect of emotions on the typing rhythm.

2.3. Features extracted from a user's particular typing rhythm

As a biometric technique, keystroke dynamics exploits features from a user's particular typing rhythm. The most common features used by researchers are timing features and frequency features. Timing features are based on calculations of single or multiple keys (called N-graph), and they are measured in milliseconds. In Figure 1, we summarize the commonly used features in keystroke dynamics related studies.

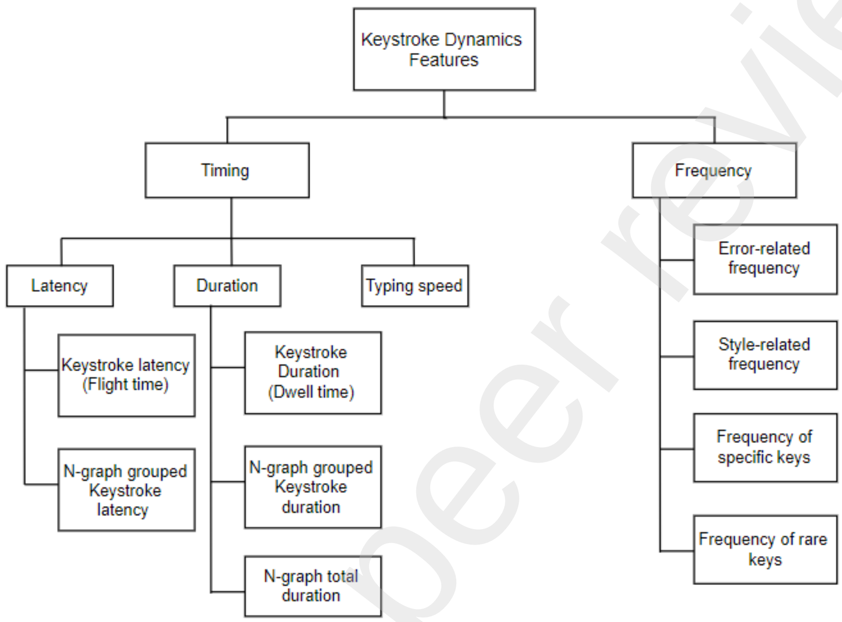


Figure 1: Keystroke Dynamics features

Keystroke dynamics consists in capturing the exact timestamp of each key pressed (Key Down) and key released (Key Up). The timestamp is the number of milliseconds that have elapsed since January 1, 1970 (midnight UTC/GMT). Based on the key events timestamps, multiple features can be extracted.

We explain in the following the common features used:

- **Dwell time:** This feature is also called keystroke duration or hold time. It represents the time between the pressing and releasing of a single key. In other words, it is the time a key is pressed. In Figure 2, DT1 is the dwell time and it is the time between KD1 and KU1 (30 - 10 = 20ms).

- **Flight time:** Also called latency or inter-key time. It is the time interval between a key release and a key press of the following key. In Figure 2, FT1 is the flight time and it is the time between KD2 and KU1 (40 - 30 = 10ms). It should be noted that this feature can also hold negative values. This happens when a key has not yet been released and a new one is pressed (FT2 = KD3 - KU2 = -10 < 0).

- **N-graphs:** It is the time between n keys up, keys down, or a combination of both events. Other than single key features, the most commonly used graphs are digraphs and trigraphs. Di-graphs groups two consecutive keystrokes, while trigraphs groups three. As an example, the word 'happy' has four di-graphs ('ha', 'ap', 'pp', 'py') and trigraphs 3 ('hap', 'app', 'ppy'). Dig1 and Trig1 shown in Figure 2 are examples of digraphs and trigraphs, respectively.

- **Frequency features:** This focuses on the use of certain keys while typing and calculates the frequency of their occurrence. Example of these keys are: error related keys (backspace and delete key frequency), style related keys (Capitalization frequency), specific keys (enter or space-bar keys frequency), rare keys (frequency of rare consonant or vowel keys).
- **Typing speed:** As presented by Eq. 1, it is the total number of keystroke or words per unit of time (usually measured in seconds).

$$Typing\ speed = \frac{KD_1 - KU_n}{Total\ number\ of\ characters\ typed} \quad (1)$$

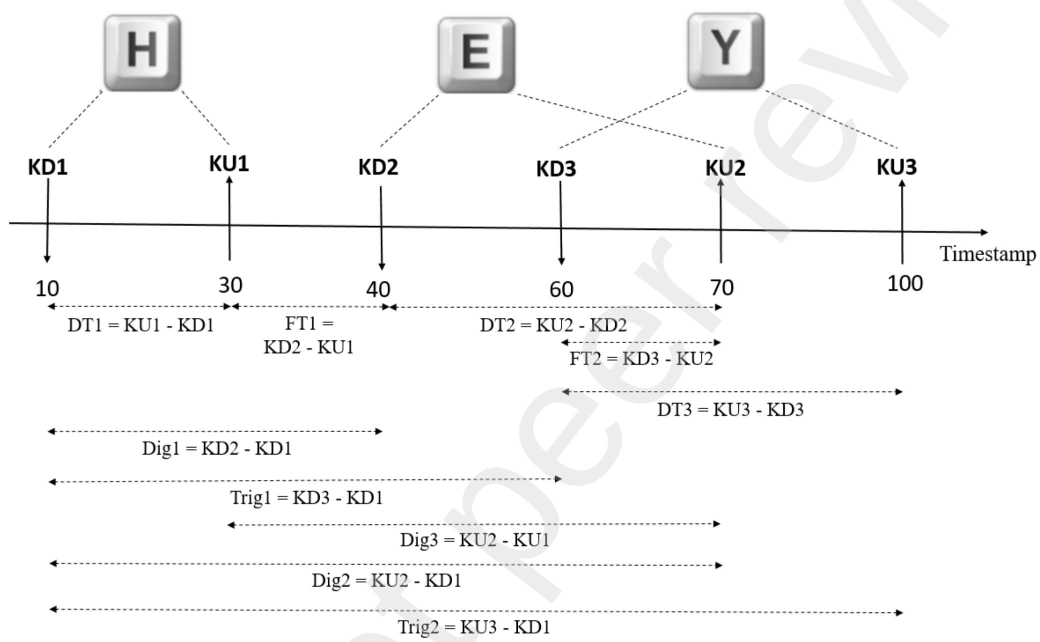


Figure 2: Keystroke Dynamics features calculations

2.4. Datasets modeling

- Emotion models:

  In order to explain the nature of emotions and characterize them, psychologists developed thee major emotion models: discrete, dimensional, and cognitive appraisal model.

  - The Discrete Model: groups emotions into discrete categories based on their similarities. This set of discrete categories of emotions is described as basic or fundamental emotions and is considered to be universal across humans [9], [32], and [33]. In computational linguistics research, the theory that is often applied is Ekman's theory [32]. Ekman classifies emotions into six classes called basic emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise).
  - The Dimensional Model: each dimension defines regions within which distinct emotions can be located. This dimensional perspective describes emotions in terms of: (i) two dimensional [34]: Valence (Denotes the polarity of emotion by describing how positive or negative is the feeling.) and Arousal (Denotes the intensity of emotion and refers to the sense of energy or the degree of activation of an individual), (ii) three dimensional: includes again the dimensions of valence and arousal, and add the dimension of dominance (also called control or power), which shows how strong an emotion is [35], or (iii) four dimensional: these dimensions are evaluation-pleasantness, potency-control, activation-arousal, and unpredictability [36].

- The Cognitive Appraisal Model: known as the OCC (Ortony, Clore, and Collins) theory [37], and claim that there is an essential and profound cognitive basis for emotions.

- Emotion elicitation method:

  In order to collect emotionally-labelled data, researchers opt for different elicitation methods to induce a specific emotion in the participants. Emotions can be induced by video clips, pictures, task difficulty, time limits, etc. However, some studies choose to gather data without inducing any emotion in the user following an experience-sampling approach. This approach relies on the self-reports of the participants to reveal their current emotion, i.e, users have to answer questionnaires, sometimes with Likert scales, or use a *free-text* box to express their currently felt emotion.

- Task for data collection:

  Keystroke dynamics has been explored and analysed by using either *Fixed-texts*, *free-texts*, or both. The main advantage of using *Fixed-text*s is that the "cognitive skills" of users will not influence the typing process; meaning that users don't have to think of something before typing. Even though *fixed-text* studies yield better results than *free-text* ones [38], some researchers use *free-texts* as a way for a closer real-world scenario.

Table 1 is an updated version of a table from our paper [39] in which we summarize the dataset parameters used in the literature to model the dataset.

Table 1: Different datasets parameters used in the literature

Emotion model	Number of emotions	Number of users	Recording tool	Data labeling	Emotion induction	Task for data collection	Ref.
Discrete	4	50	Desktop App	Self reports	-	-	[40]
	15	12	Desktop App	Self reports	-	Fixed/Free	[41]
	7	9	Desktop App	Self reports	-	Free	[42]
	7	25	Desktop App	Self reports	Recall tasks	Fixed/Free	[43]
	1	9	-	-	-	Free	[44]
Dimensional	5	35	Desktop App	Self reports	Videos/pictures	Free	[45]
	-	14-27	Desktop App	Self reports	Task difficulty	Free	[26]
	-	27	Desktop App	Self reports	Pictures	Fixed	[46]
	-	152	Mobile App	Self reports	3-time limits	Free	[47]
	14	64	Web App	Self reports	Videos/Tasks	-	[28]

2.5. *Classification techniques and best accuracy rates achieved while studying emotion recognition through Keystroke Dynamics*

In [39], we conducted a systematic literature review where we considered and analyzed 10 papers that investigate emotion recognition through keystroke dynamics. Table 2, taken from our paper [39], outlines the techniques used for emotion recognition and the accuracy rates in 10 different studies.

3. Building a Dataset on Keystroke Dynamics

Due to the scarcity of datasets related to our purpose, and the fact that performance in keystroke dynamics is highly dependent on the dataset, we developed a dynamic web application, to construct a new dataset. Figure 3 summarizes the main steps that we undertake throughout the present research work.

3.1. *Experimental Set-up*

In this section, we give details of the emotion model, emotion elicitation method, and the tasks to collect the data.

Table 2: Accuracy rates and techniques used for classification

Approaches	Techniques / Methods	Best accuracy rates	Ref.
Machine learning	ANN and SVM	91.24% accuracy and 4.35% FP rate using SVM (fright)	[40]
	Decision tree	77% to 88% for hesitance, nervousness, relaxation, sadness, and tiredness	[41]
	Logistic regression, SVM, Nearest neighbor, C4.5, and Random Forest	84% for arousal using KNN and 83% for valence using KNN	[28]
	Decision tree, Neural Networks, K Nearest Neighbors, Naive Bayes, AdaBoost, Rotation Forest, and Bayesian Network	81.25% using AdaBoost for fear	[42]
	C4.5, tree J48, Random Forest, SVM, and Naive Bayes	80,6% for negative and positive arousal using Random Forest	[26]
	Simple logistics, Sequential minimal optimization (SMO), multi-layer perceptron, Random Tree, J48, and BF tree	Between 70% and 80%for *fixed-text* / 60% to 82% for *free-text*	[43]
	KNN, KStar, Random Committee, Random Forest, and Bounded K-means Clustering	69% using Random Comittee for positive and negative emotions	[45]
Statistical	Two-way 3(Valence: negative, neutral, positive) * 3(Arousal: negative, neutral, positive) ANOVA	Significant (p<0.001)	[46]
	linear regression model	Significant (P<0.01)	[47]
	Logistic regression	60% to 88% accuracy with mean 72%	[44]

3.1.1. Hybrid Emotion Model

Since there is no clear evidence on which emotion model yields better accuracy results, we opt for a hybrid model which plots four discrete emotions in the two-dimensional model. This model is an adjusted model of the discrete model presented in our previous paper [39]. We choose this model for three main reasons:

- First, we select three emotions (sad, happy, angry) from the discrete model because their terminology is already familiar to participants due to their prevalence in the everyday language;
- Second, discrete emotions are characterized by stable patterns of triggers and represent a unique mechanism that causes a unique mental state with unique measurable outcomes [48];
- Third, we add the "calm" emotion to fit the set of the emotions to the two-dimensional model in a way that we reflect the idea of High/Low Arousal and Positive/Negative Valence.

Thus, "Angry" will be placed in the area of High Arousal and Negative Valence, "Happy" in High Arousal Positive Valence, "Sad" in Negative Valence and Low Arousal, "Calm" in Positive Valence and Low Arousal, and "Neutral" in Medium Arousal and Zero Valence.

3.1.2. Emotion Elicitation Method

Unlike studies in [40, 41, 42, 43], researchers relied on participants' self-reports by asking users questions about their current emotions. In our work, we choose to induce emotions in participants through video clips. We adopt this method for two main reasons; first, labeling emotions based on participants' own interpretations of their emotional state could be misleading and not true. For instance, a person who is sad may disregard this feeling and reports that he is in an other emotional state. It may also be strange to ask someone about his current emotion. Second, emerging studies, like in [49, 50], confirm that videos are one of the most effective methods of emotion induction.

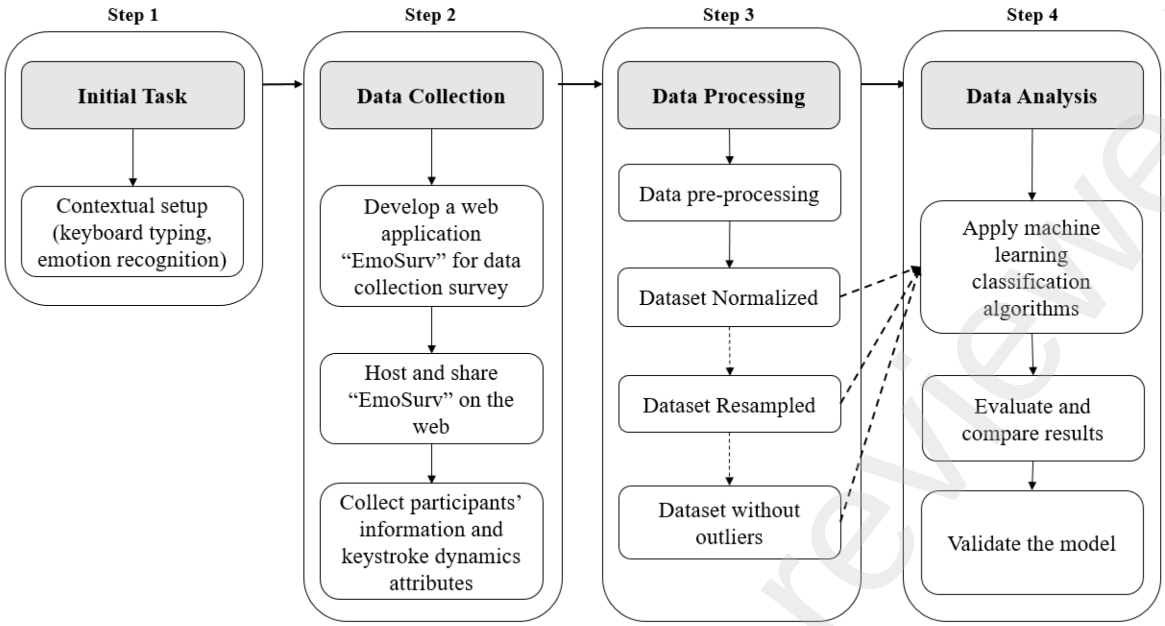


Figure 3: Method's flowchart to generate a model from data related to keystroke dynamics and emotion states

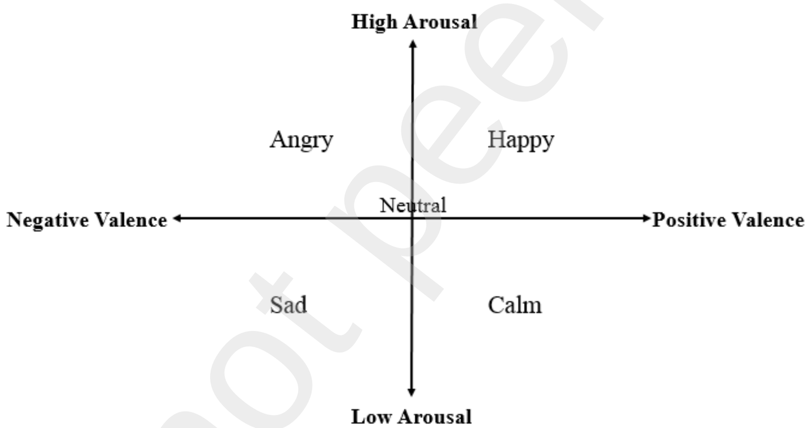


Figure 4: Hybrid emotion model

We carefully select the emotion-eliciting videos from the study in [49]. In fact, the researchers spent four decades of investigation and focused on testing the efficacy of a set of 15 evocative videos in eliciting emotions on 784 adults. We describe in Table 3 the chosen films for each emotion.

In order to make sure that, while capturing the data, the participants are engaged and watch the whole video, we include an accuracy question related to each video in a way that every participant is asked to respond to an MCQ (Multiple Choice Question) after viewing a video. In the case of a wrong answer, data related to that video will be removed in the cleaning phase.

3.1.3. Task Scenarios for Data Collection

To perform the data acquisition phase, participants are asked to type on their keyboards free and *fixed-texts* before and after watching the emotion-eliciting videos. For *fixed-texts* typing, the participant can go to the next page only if the similarity rate between the given text and the typed text is higher than 90%. Here, the participant is given a short and simple text related to the previously watched video. While preparing the *fixed-text*s, we use specific terms taken from [51]. Each term is a characterization of an emotion. As for the *free-texts* typing, the

Table 3: Selected video clips associated to each emotion in EmoSurv

Emotion	Video name	Source	Description
Angry	Once Were Warriors	Movie	A 1 munite and 6 seconds video about a drunk man aggressively violating his wife.
Sad	The Champ	Movie	A 2 munites and 51 seconds video showing a sad young child crying for losing his beloved father.
Calm	Nature clip	YouTube video	A 7 minutes video of relaxing music and beautiful nature.
Happy	Funny cats and babies	YouTube video (joined video)	A 2 minutes and 19 seconds video composed of a sequence of funny cats and babies' reactions.

"Next" button that takes to the following task is activated only if the participant types more than 40 characters.

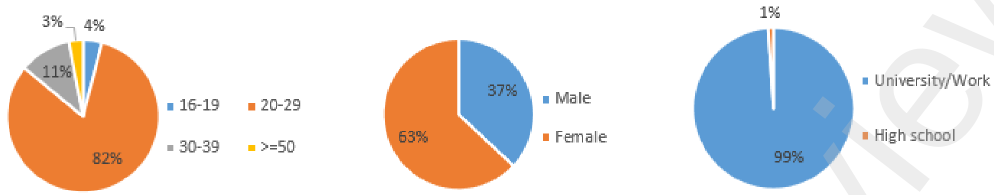
3.2. Building the Dataset

One of the cornerstones upon which the present work rests on, is the set-up and development of the data collection software to build the dataset. This phase is detailed in our paper [39]. Table 4, presents the elements of the extracted datasets.

Table 4: Raw datasets description

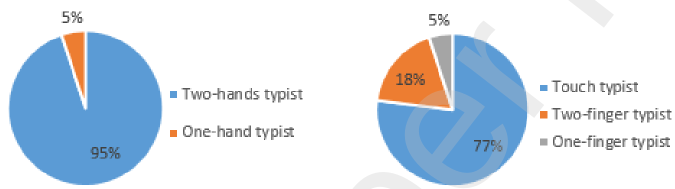
Dataset name	Elements	Description
User dataset	User ID	An ID is allocated for each user.
	Demographic information	Age range, gender, status (Student / professional), degree (high school / university), country
	Typing-related information	Typist type 1(one-hand-typist / two-hand-typist), typist type 2 (one-finger-typist / two-finger-typist / finger-typist), average time spent using computer keyboard per day.
Fixed-text typing dataset	User ID	An ID is allocated for each user.
	Emotion index	"N" for Neutral, "H" for Happy, "A" for Angry, "S" for Sad, and "C" for Calm
	Index	Incremented by 1 every-time a key event takes place.
	Key code	A code that represents a keypress.
	time related features	Key Down and Key Up timestamps
	Answer	log "R" when accuracy question answered right, and "W" when accuracy question answered wrong.
Free-text typing dataset	Same as *fixed-text* typing.	Same as *fixed-text* typing.
Frequency dataset	User ID	An ID is allocated for each user.
	Text index	"FT" for *free-text* / "FI" *fixed-text*
	Emotion index	"N" for Neutral, "H" for Happy, "A" for Angry, "S" for Sad, and "C" for Calm
	DelFreq	Takes the number of how many times the delete key is pressed.
	LeftFreq	Takes the number of how many times the back arrow key is pressed.
	Typing speed	Calculated only for *fixed-text* typing as the number of keys pressed over the time spent from the first key pressed to the last key released.

Our web application "EmoSurv" is deployed on the 14$^{\text{th}}$ of January, 2020. It is shared to participants on the day after; from the 15th of January to the 15th of March via emails, Facebook, and word of mouth. By the time we generate the dataset from EmoSurv, we count 124 participants who have visited EmoSurv and logged information. Figure 5 shows the demography of the participants and Figure 6 shows the category of the participants based on the number of hands and fingers used to type on a keyboard. This data is not analyzed in this paper, but it will be considered as a larger and stratified samples on futur works.



(a) Age range *(b)* Gender *(c)* Educational level

Figure 5: Users taxonomy based on their age range, gender, and educational level



(a) Number of hands used to type *(b)* Number of fingers used to type

Figure 6: Users taxonomy based on the number of hands and fingers used to type

The difference regarding the environment in which the samples are gathered may affect the results. Here, we mean by environment the type of computer used (desktop or laptop), the place where the survey is taken (home, office, etc.), the time, or any other external factor. However, it is worth to mention here that one of the important aspects of our research work is to collect keystroke dynamics data in an environment as close as possible to real life scenarios and without any form of intervention.

3.3. Data Capturing and Feature Extraction

In a very first step, data collected from EmoSurv application is exported to CSV files. As detailed in Table 4, three categories of information are extracted:

- Keystroke features: divided into timing and frequency related features and stored in two different files;
- Various information about the users stored in "users dataset";
- Emotional state: which is the label for the extracted Keystroke dynamics data.

The raw keystroke data consists of the key code, the timestamp of key press (Key Down), and the timestamp of key release (Key Up). To generate features, we used graphs (a single key event), digraphs (a combination of 2 consecutive keystroke events) and trigraphs (a combination of 3 consecutive keystroke events). As for frequency and speed features, the values are calculated per task, where the task is considered as the whole paragraph typed by the user in a specific emotional state. Table 5 describes the 10 keystroke attributes that are extracted during the data collection process. All the features are calculated for the fixed and *free-text*, before and after the emotion elicitation.

Special consideration: In some cases, a user keeps pressing on a key for a longer time, causing to write that key letter multiple times ('hhhhhhhhhhh'). In this case, if one press and release action results in generating more than 5 times the same letter, then all the logs for that key are aggregated in one row. For the values of key up and down timestamps, they take the first and last timestamps captured, respectively.

Table 5: Description of coded keystroke attributes with the number of associated graphs

Features	Notation as in Figure 2	Description	Graphs
D1U1	DT1	Time between first key down and first key up	1
D1U2	Dig2	Time between first key down and second key up	2
D1D2	Dig1	Time between first key down and second key down	2
U1D2	FT1 / FT2	Time between first key up and second key down	2
U1U2	Dig3	Time between first key up and second key up	2
D1U3	Trig2	Time between first key down and third key up	3
D1D3	Trig1	Time between first key down and third key down	3
DelFreq	-	Relative frequency of delete key	NA
LeftFreq	-	Relative frequency of backspace key	NA
Typing speed	-	Number of key pressed in each task the time spent from the first key pressed to the last key released (in the same task).	NA

3.3.1. Data Preprocessing

The preprocessing phase is conducted using Python, through the Anaconda navigator. To deal with missing values, we go with the strategy of imputation by computing the overall mean. In our dataset, missing values can be generated when a user types the last key; the features D1U2, D1D2, U1D2, and U1U2 can not be calculated. For instance, D1U1 is the difference between the i key down and the i+1 key up, where in the case of i=n (n being the last key in the task), the key i+1 does not exist. Same applies for D1D2, U1D2, and U1U2. In such a case, we have one missing value for each of the 4 features in each task. As for D1U3, and D1D3, we have 2 missing values for each of the 2 features in each task.

We also conducted additional data cleaning by removing data entries related to mobile phones or tablets (virtual keyboard), incomplete data, and submissions that are recognized as having wrong answers for the accuracy answers.

3.3.2. Size and Description of the Obtained Datasets

Throughout the data preprocessing phase, the size of the datasets decreases, and the number of participants is reduced from 124 to 88. Figure 7 and Figure 8 depict the size evolution of *fixed-text* and *free-text* datasets, respectively.

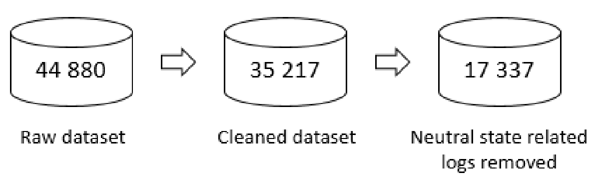


Figure 7: Size evolution, as per number of rows, of *fixed-text* dataset along the data preprocessing phase

Figure 9 presents the number of instances collected for each emotion class (Happy, Calm, Angry, and Sad).

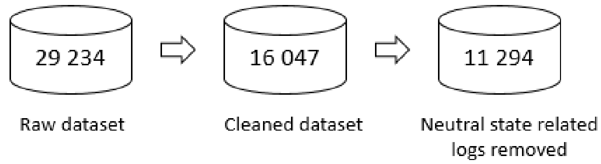


Figure 8: Size evolution, as per number of rows, of *free-text* dataset along the data preprocessing phase

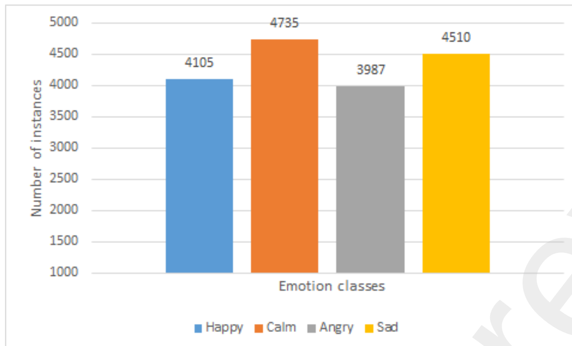


Figure 9: Fixed-text dataset: Number of instances in each class

4. Data Analysis and Results Discussion on Models Validation

Our methodology starts with generating different datasets from the *Fixed-text* and the *Free-text* datasets. Then we train, build, and compare the models generated from those datasets. To this end, we begin by applying techniques like standardization (resulting in the Stand-Dataset), resampling (resulting in the Res-Dataset), and outlier removal (resulting in the Out-Dataset). Second, we examine each dataset by its own; We split data using ten-fold cross validation. Third, we train the data using supervised machine learning algorithms.

The four datasets are used separately to generate different models. By following this methodology, we investigate how the standardization technique, resampling filter, and removing the outliers impact the quality of predicted models. We summarize in Figure 10 the methodology adopted: (i) examin the extent to which the Raw Datasets, Stand-Datasets, Res-Datasets, Out-Dataset, and Freq-Datasets are reliable to recognize emotions, (ii) compare the Raw Dataset and the Stand-Dataset to reveal the impact of user-specific standardization on the quality of the predicted model, (iii) compare the Stand-Dataset and Res-Dataset to reveal the impact of applying the resampling filter on the quality of the predicted model, (iv) compare the Res-Dataset and Out-Dataset to reveal the impact of removing the outliers on the quality of the predicted model, and (v) compare the models in terms of the type of text, which could be either fixed or free.

4.1. Generated Datasets

Having the cleaned datasets (*free-text* and *fixed-text*) in hand , we generate three versions of datasets. These are drawn in Figure 11 and detailed in the following sections. Stand-Dataset is generated after applying user-specific data standardization on the Raw Dataset. Res-Dataset is obtained after applying the resampling filter on the Stand-Dataset. After removing outliers from the Res-Dataset, we obtain the Out-Dataset.

4.1.1. Raw Dataset

This dataset is the cleaned version of the raw dataset initially obtained. It is generated after handling missing values and applying the additional data cleaning steps (explained in Section 3.3.1).

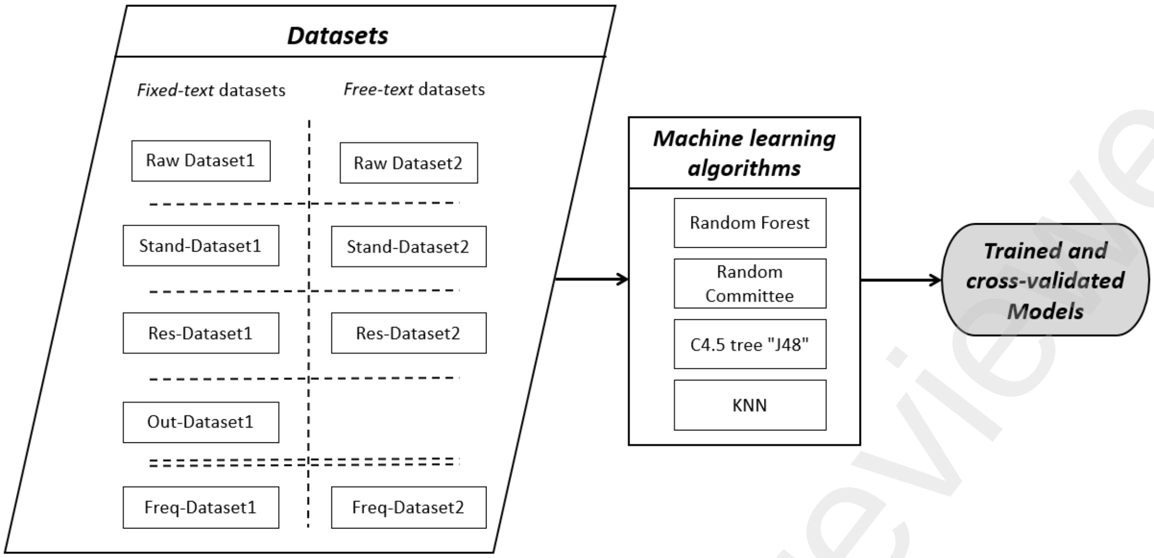


Figure 10: Adopted methodology for data analysis

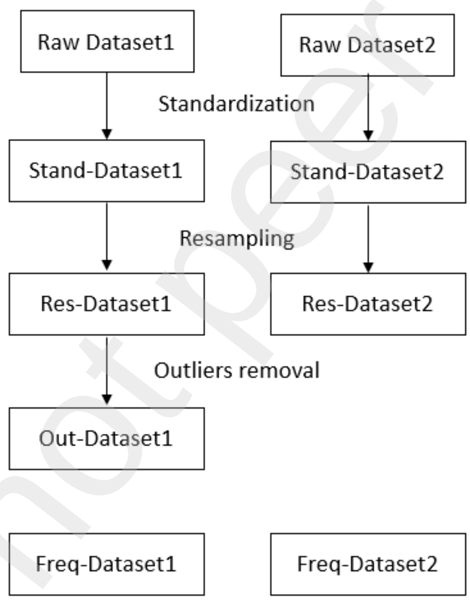


Figure 11: Generated datasets

4.1.2. Stand-Dataset: The Standardized Dataset

Numerical data extracted from the typed texts may have very different ranges, and direct comparison is often not meaningful. Feature scaling is a way to bring all data into a similar range for a more useful measuring. For this reason, we adopt the "Z-score" (or "standard score") to standardize the data.

Particularly in this research, we calculate standardized values for each user separately. User-specific standardization is implemented to insure that users' performances and abilities are taken into consideration. The formula applied for "standard score" is given by Eq. (2):

$$Standard\ score\ (i) = \frac{X_i - \mu_i}{\sigma_i}, \tag{2}$$

where μ is the mean of one feature instances of each user in one specific task. And σ is the standard deviation of one feature instances of each user in one specific task.

Figure 12 presents the distribution of "D1U1" feature as a Gaussian distribution.

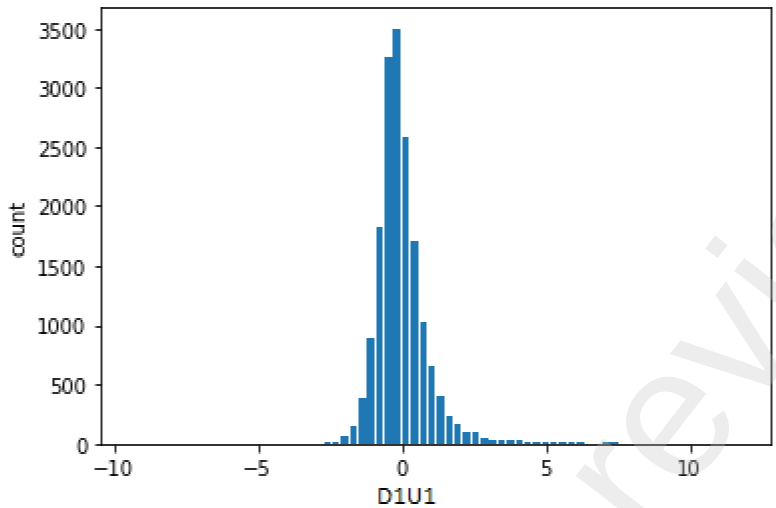


Figure 12: Distribution of "D1U1" feature

4.1.3. Res-Dataset: The Resampled Dataset

This phase is performed using the Graphical User Interface (GUI) of the WEKA tool.

Since filtering methods were not rigorously studied in previous works that hold the same objective as ours, we investigate on the resampling filter from the WEKA tool. The resampling filter in WEKA tool produces a random subsample of a dataset using either sampling with replacement or without replacement [52]. The default option, which is "resampling with replacement" is tested. And since we have a nominal class attribute, the supervised version of the filter is applied.

4.1.4. Out-Dataset: Dataset with Outliers Removed

In view of the particularity of the data, we are skeptical about the effect of outliers when building the model. We believe that a decrease in model accuracy may happen due to outlier values. Therefore, we generate the Out-Dataset to evaluate whether or not the process of removing outlier values would decrease the accuracy of the models. The Out-Dataset is obtained after removing outliers from the standardized and resampled dataset.

In our research work, the examination of the outliers is performed by feature, where each feature is handled separately. We conduct the outliers detection and removal using Python.

***Once detected, outliers are removed using 3 sigma (why) as outlier threshold [53] (page 19) and the obtained dataset is saved as the Out-Dataset. The size of the *fixed-text* dataset decreased from 17 338 to 16 507 instances after removing the outliers. ***

4.1.5. Freq-Dataset: Dataset Containing Frequency Features

Along with the timing features, frequency features are also collected. In this section, we consider these frequency features namely: Delete Frequency, Left Arrow Frequency, and the Typing Speed. Resampling filter is also applied to this version of dataset.

4.2. Models Building

This section introduces the methods used for the evaluation of the methods used and our research methodology.

Evaluation Criteria

There are several evaluation criteria used to evaluate and measure the performance of the obtained affective model:

- F-Score: it is the Harmonic Mean between precision and recall. The range for F-Score is [0,1]. It is a reflection of how precise (how many instances it classifies correctly) and how robust (it does not miss a significant number of instances) the classifier is. We use F-score metric when models are built for each emotion separately; in this case, the classes are imbalanced.

  High precision but lower recall, gives an extremely accurate, but it then misses a large number of instances that are difficult to classify. The greater the F-Score, the better is the performance of our model. Mathematically, it can be expressed as shown in Eq. (3):

$$F-Score = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Precision}} \tag{3}$$

- Kappa: It represents the extent to which the data is a correct representations of the variables measured [41].

- Accuracy: Is used to measure how often the algorithm classifies a data point correctly [40]. We use this measure when we are evaluating the type of text used (free or fixed text), which is a similar class distribution. It is calculated using the following Eq. (4):

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made} \tag{4}$$

- Computational time: called also process time. It is the time that is taken for each classification from the beginning to the end. This criterion is mostly related to the nature of the machine learning method and the amount of data for training and testing iterations. This parameter is measured in seconds, and lower values reflect faster processing time.

Evaluating the Machine Learning Models

After discussing the different ways to consider the data (Raw Dataset, Stand-Dataset, Res-Dataset, Out-Dataset, Freq-Dataset), we focus now on how to generate and test the predictive models.

4.2.1. Raw Dataset Based Model

Here, we train the models using the timing features from the Raw Dataset. Figure 13 and Figure 14 display F-score values obtained after applying J48, Random Forest, Random Committee, and KNN on the *fixed-text* and *free-text* Raw Datasets, respectively. The highest F-score values achieved from this model are 0.395 and 0.474 after applying the Random Forest algorithm on the *fixed-text* and *free-text* datasets, respectively. As for accuracy rates, Table 6 shows that the best accuracy rate is 37.5808 % with kappa statistic equal to 0.136.

These values are very low, hence we deduce that using the Raw Dataset to generate predictive models is not reliable.

4.2.2. Stand-Dataset Based Model

Here, we train the models using the timing features from the Stand-Dataset. This dataset is obtained after applying user-specific standardization technique explained in section 4.1.2.

Figure 15 and Figure 16 below display F-score values using *fixed-text* and *free-text* datasets, respectively. For the *fixed-text* dataset, Figure 15 shows F-score values ranging between 0.679 and 0.815, with 0.815 being the highest

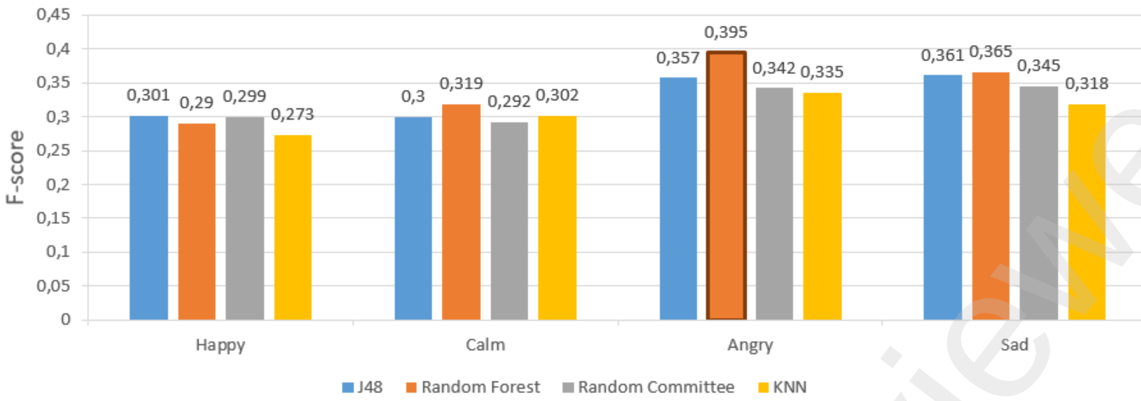


Figure 13: F-score results obtained using the Raw Dataset (*fixed-text*)

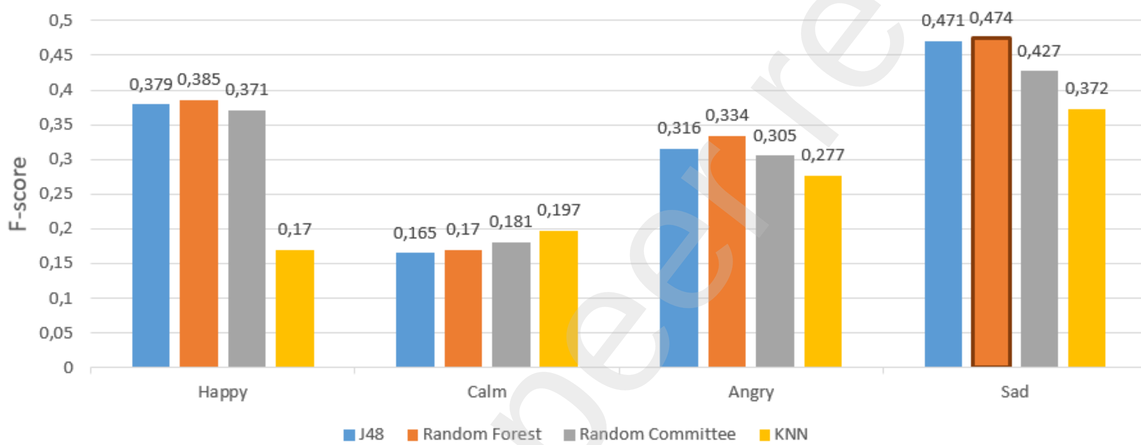


Figure 14: F-score results obtained using the Raw Dataset (*free-text*)

Table 6: Comparing results of the different machine learning algorithms applied to the *fixed-text* and *free-text* datasets (Raw Dataset)

	Fixed-text dataset results			*Free-text* dataset results		
	Comp. time	Accuracy	Kappa	Comp. time	Accuracy	Kappa
J48	4.4 seconds	32.9937 %	0.1038	1.74 seconds	37.0229 %	0.127
Random Forest	34.14 seconds	**34.2325 %**	**0.1206**	20.52 seconds	**37.5808 %**	**0.136**
Random Committee	11.76 seconds	31.8708 %	0.0898	3.02 seconds	34.6144 %	0.1001
KNN	**0.03 seconds**	30.6842 %	0.074	**0 seconds**	30.3905 %	0.0544

value captured for "Calm" emotion using Random Forest algorithm. As for the results from *free-text* dataset, F-score values range between 0.642 and 0.832, with 0.832 being the highest value captured for "Happy" emotion using Random Forest and J48 algorithms (See Figure 16).

Table 7 exhibits the accuracy along with the Kappa statistic and the computational time for both *fixed-text* and *free-text* datasets. For instance, for the *fixed-text* the accuracy rates vary between 72.6727 % and 76.4818 %, with Random Forest achieving the highest percentage with a Kappa of 0.6855. However, this algorithm takes the longest time (27.17 seconds) to build the model compared to the other algorithms (between 0.03 to 4.79 seconds).

Even though no filtering or further processing is applied to this version of the dataset, we obtained considerably high F-scores, and accuracy rates. We also notice that values obtained from both datasets (*fixed-text* and *free-text*) are very "similar" (almost the same rates).

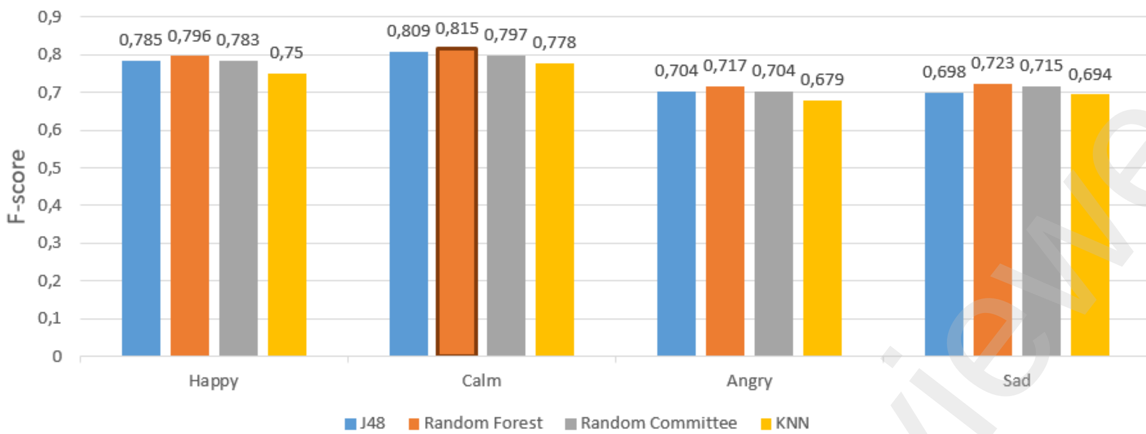


Figure 15: F-score results obtained on each emotion after training classifiers using the *fixed-text* dataset (Stand-Dataset)

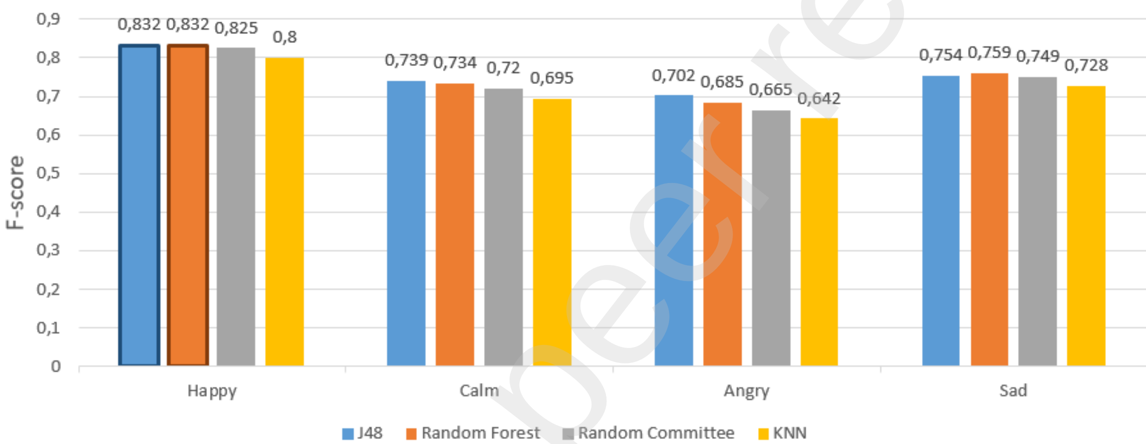


Figure 16: F-score results obtained on each emotion after training classifiers using the *free-text* dataset (Stand-Dataset)

Table 7: Comparing results of the different machine learning algorithms applied to the *fixed-text* and *free-text* datasets (Stand-Dataset)

	Fixed-text dataset results			*Free-text* dataset results		
	Comp. time	Accuracy	Kappa	Comp. time	Accuracy	Kappa
J48	2.6 seconds	75.2294 %	0.669	0.66 seconds	**76.5435** %	**0.6827**
Random Forest	27.17 seconds	**76.4818 %**	**0.6855**	17.68 seconds	76.238 %	0.6773
Random Committee	4.79 seconds	75.1428 %	0.6677	2.77 seconds	75.0876 %	0.6613
KNN	**0.03 seconds**	72.6727 %	0.6351	**0 seconds**	72.706 %	0.629

4.2.3. Res-Dataset Based Model

Here, we train the models using the timing features from the Res-Dataset. This dataset is obtained after applying the resampling filter.

For the *fixed-text* dataset, Figure 17 shows F-score values that range between 0.806 and 0.951, with a highest value of 0.951 being achieved for "Anger" emotion using Random Committee.

As for the results from *free-text* dataset, F-score values range between 0.802 and 0.931, with 0.931 being the highest value captured for "Happy" emotion using Random Forest algorithm (Figure 18).

As in the previous model, Table 8 indicates that the values obtained from both datasets are very "similar" (almost the same rates). For instance, for the *fixed-text* the accuracy rates vary from 84.2846 % up to 93.922 % for Random Committee.

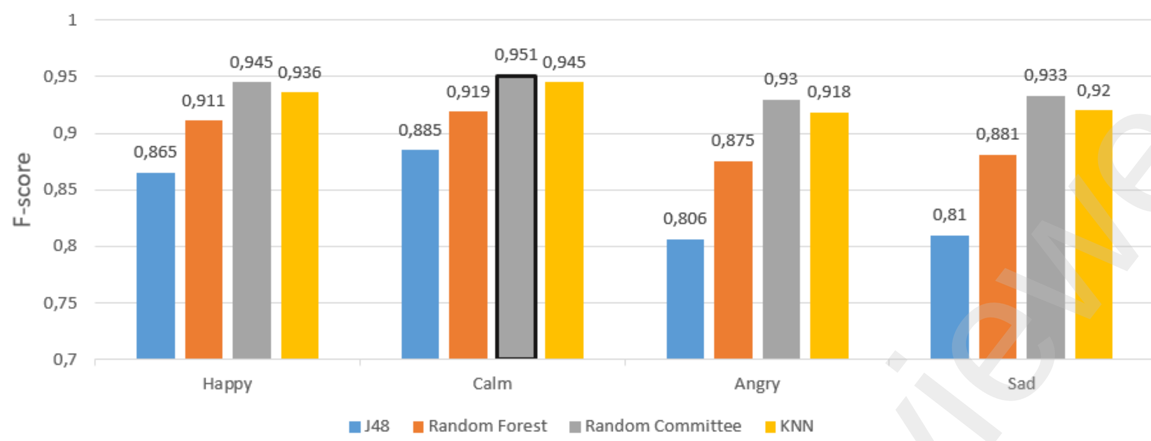


Figure 17: F-score results obtained on each emotion after training classifiers using the *fixed-text* dataset (Res-Dataset)

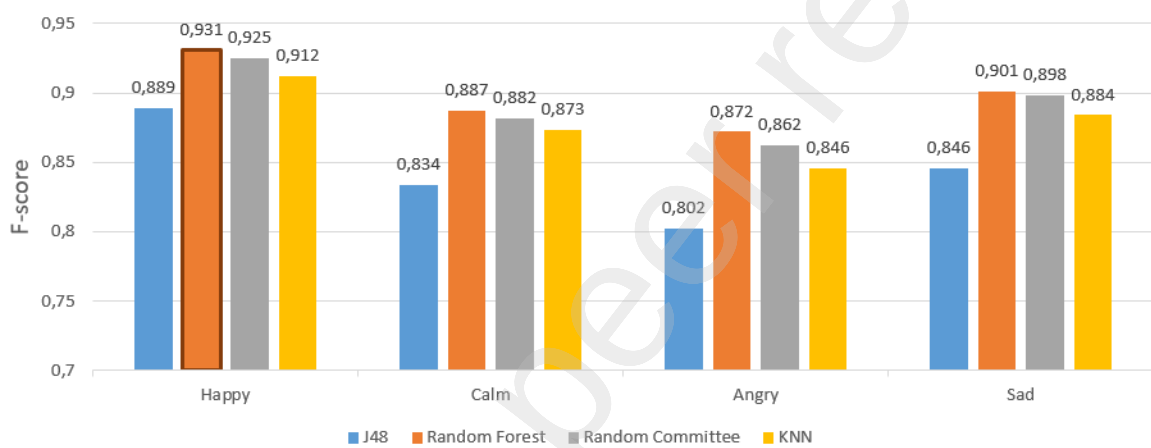


Figure 18: F-score results obtained on each emotion after training classifiers using the *free-text* dataset (Res-Dataset)

Table 8: Comparing results of the different machine learning algorithms applied to the *fixed-text* and *free-text* datasets (Res-Dataset)

	Fixed-text dataset results			*Free-text* dataset results		
	Comp. time	Accuracy	Kappa	Comp. time	Accuracy	Kappa
J48	1.26 seconds	84.2846 %	0.7901	0.59 seconds	84.9016 %	0.7945
Random Forest	21.99 seconds	89.7443%	0.863	14.38 seconds	**90.204%**	**0.8667**
Random Committee	3.07 seconds	**93.922 %**	**0.9197**	1.94 seconds	89.6288%	0.8588
KNN	**0.01 seconds**	93.0282 %	0.9069	**0 seconds**	88.3167%	0.8411

4.3. Out-Dataset Based Model

Here, we train the models using the timing features from the Out-Dataset. This dataset is obtained after removing outliers from the Res-Dataset.

Figure 19 presents F-score values achieved using the *fixed-text* dataset. These values range between 0.805 and 0.919, with 0.919 being the highest value attained for "Calm" emotion using Random Forest algorithm.

In this step, we notice that F-score results does not improve when we generate the Out-Dataset based model using *fixed-text* dataset; a very slight decrease in these values is noted. Thus, we do not proceed with generating the Out-Dataset based model for the *free-text* dataset.

From Table 9, we can infer that all of the algorithms perform well with an accuracy rate varying between 84.0422 % (Kappa equal to 0.7869) while using J48 and 89.795 % (Kappa equal to 0.8634) while using Random Committee.

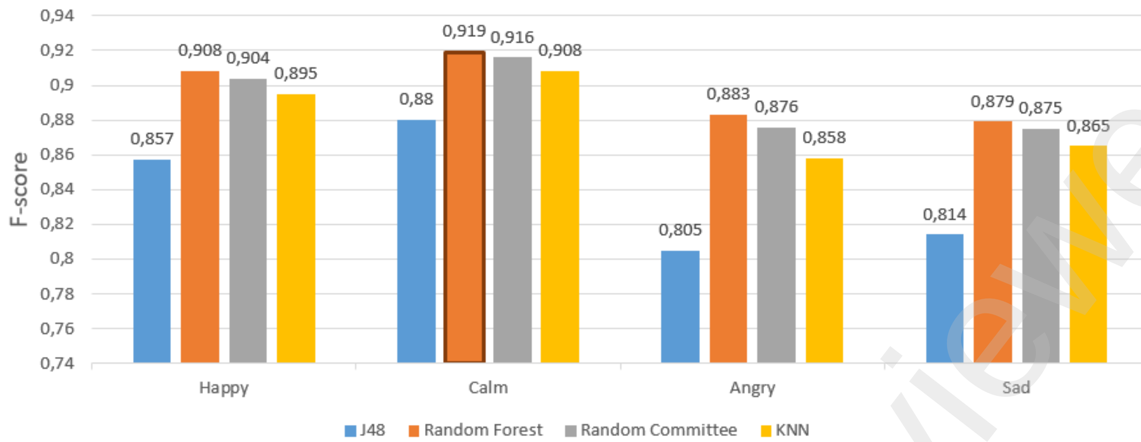


Figure 19: F-score results obtained on each emotion after training classifiers using the *fixed-text* dataset (Out-Dataset)

Table 9: Comparing results of the different machine learning algorithms applied to the *fixed-text* and *free-text* datasets (Out-Dataset)

Algorithm	Comp. time	Accuracy	Kappa
Tree J48	1.59 seconds	84.0422 %	0.7869
Random Forest	20.48 seconds	**89.7795 %**	**0.8634**
Random Committee	3.44 seconds	89.313 %	0.8572
KNN	**0.01 seconds**	88.2467 %	0.8431

4.4. Freq-Dataset Based Model

Figures 20 and Figure 21 display F-score values achieved using typing behaviors of *fixed-text* and *free-text* datasets, respectively. For the fixed-text typing, F-score values range between 0.612 and 0.769. The highest value is achieved using Random Forest algorithm while detecting "Sad" emotion. As for the free-text typing, F-score values range between 0.394 and 0.833. The highest value is achieved using Random Committee algorithm while predicting "Calm" emotion.

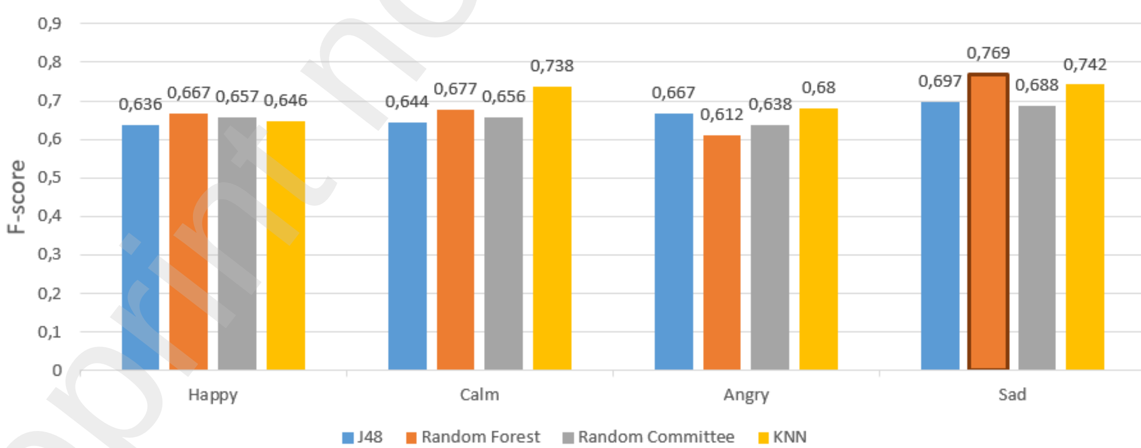


Figure 20: F-score results obtained using the frequency dataset collected from fixed-text typing

4.5. Models Comparison And Results Discussion

Figure 22 and Figure 23 draw the evolution of models' accuracy rates for the different versions of datasets generated from the *fixed-text* and the *free-text* dataset:

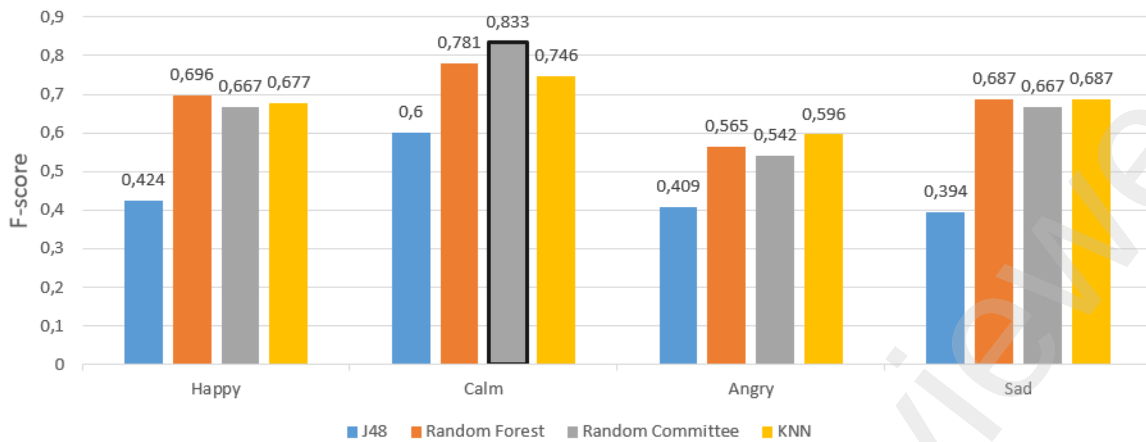
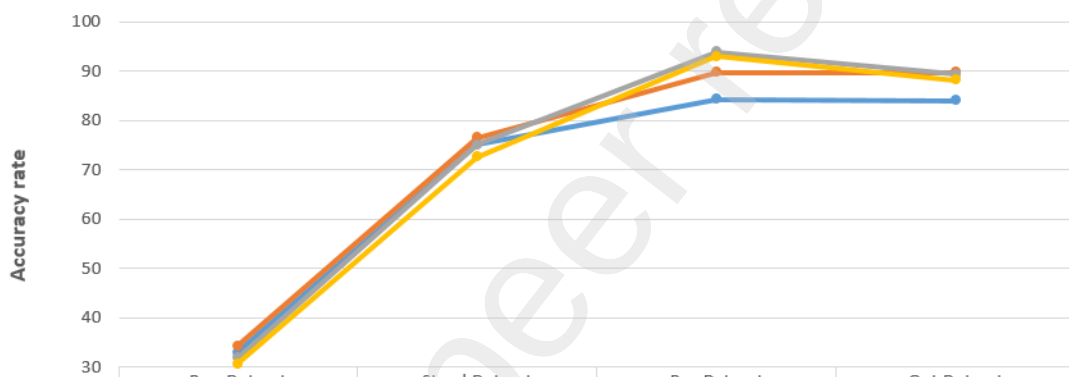


Figure 21: F-score results obtained using the frequency dataset collected from free-text typing



	Raw Dataset	Stand-Dataset	Res-Dataset	Out-Dataset
J48	32,9937	75,2294	84,2846	84,0422
Random Forest	34,2325	76,4818	89,7443	89,7795
Random Committee	31,8708	75,1248	93,922	89,313
KNN	30,6842	72,6727	93,0282	88,2467

Figure 22: Evolution of the models' accuracy rates for the different versions of datasets generated from the *fixed-text* dataset

- A remarkable increase in accuracy values is noticed when going from the Raw Dataset to the Stand-Dataset based models (i.e using the Random Forest algorithm in the *fixed-text* dataset, values increased from 34.2325 % to 76.4818 %). This proves that considering participants' variations while typing is a very important step when studying keystroke dynamics to recognize emotions. When compared to similar works [26] (the only study that considered participants' variations), we find that using the user-normalized dataset, the accuracy rate and kappa are lower than 50 % and 10 %, respectively. Therefore, applying user-specific standardization (using Z-score formula in Eq. 2) has a better effect on improving the quality of the model.

- The Res-Dataset reached higher accuracy rates (highest values achieved) compared to the Stand-Dataset (i.e using the Random Committee algorithm in the *fixed-text* dataset, values increased from 75.1248 % to 93.922 %). Thus, applying the resampling filter improves the quality of the models.

- When examining Figure 22, we notice that accuracy values slightly decrease when going from the Res-Dataset to Out-Dataset (i.e using the KNN algorithm, values decreased from 93.0282 % to 88.2467 %). This might be due to additional information that outliers can (add) when it comes to emotion recognition. And hence, we deduce that removing outliers does not improve the quality of the models.

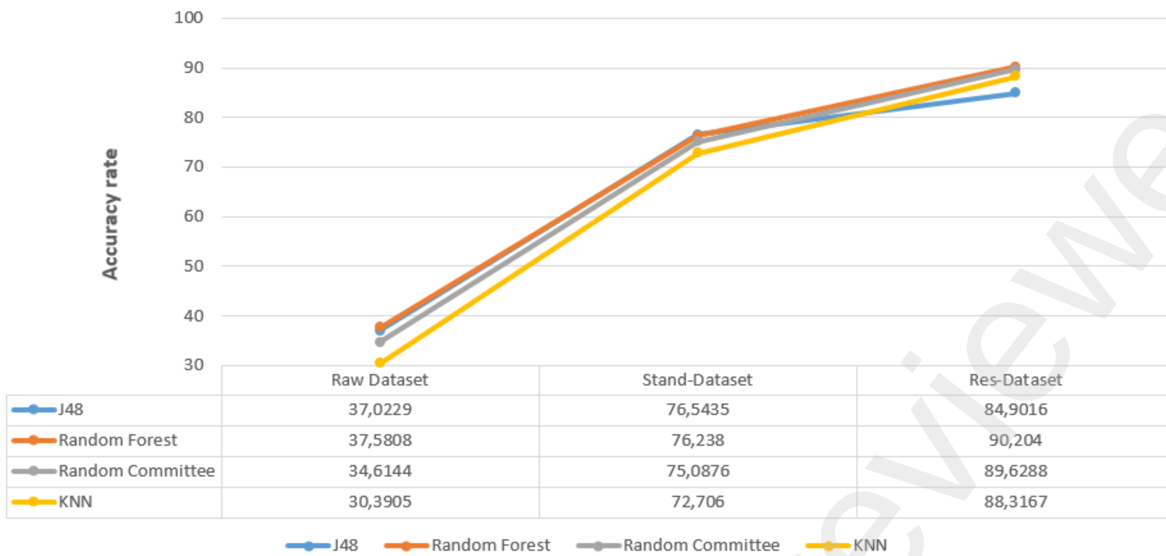


Figure 23: Evolution of the models' accuracy rates for the different versions of datasets generated from the *free-text* dataset

Table 10 presents the highest F-score values achieved by each machine learning algorithm along with the relative emotion, techniques applied, and the type of typed text. Highest F-score values are obtained using both user-specific standardization and resampling techniques. Not only fixed-text typing generates high F-score results, but also an F-score value of 0.931 is achieved using Random Forest in the case of free-text tying. Whereas in [43], for free-text typing, the success rate for the emotions "Anger", "Sadness", and "Joy" are 66 %, 60 %, and 82 % respectively.

Table 10: Best performing model for each algorithm

Algorithm	F-score	Emotion	Techniques applied	Type of text
Tree J48	0,889	Happy	User-specific standardization and re-sampling	Fixed text
Random Forest	0,931	Happy	User-specific standardization and re-sampling	Free text
Random Committee	0,951	Calm	User-specific standardization and re-sampling	Fixed text
KNN	0,945	Calm	User-specific standardization and re-sampling	Fixed text

Table 11 ranks the applied algorithms based on their achieved accuracy and kappa values. Although the four algorithms achieved high values, Random Forest algorithm often gets the best accuracy rates. However, using the Res-Dataset, from where the best models are built, the best performing algorithms are Random Forest for the *free-text* and Random Committee for *fixed-text* typing datasets.

Table 11: Ranking Algorithms based on their achieved accuracy and Kappa

Algorithm	Raw Dataset		Stand-Dataset		Res-Dataset		Out-Dataset
	Fixed-text	Free-text	Fixed-text	Free-text	Fixed-text	Free-text	Fixed-text
J48	2	2	2	1	4	4	4
Random Forest	1	1	1	2	3	**1**	1
Random Committee	3	3	3	3	**1**	3	2
KNN	4	4	4	4	2	4	3

5. Conclusion

This research work investigated a solution approach for human emotion recognition using non-intrusive interaction source which is the computer keyboard. This approach is outstanding from different aspects.

On one hand, if we are to compare our approach to others such as signal and image processing, we notice that in terms of computational time, keystroke dynamics technique incurs much less time; this validates the real-time aspect of the emotion recognition system. Furthermore, no extra hardware is required since keyboards are mostly available in all personal computers, eliminating all privacy issues and inconveniences caused by the use of cameras, microphones, or any other data acquisition tools.

On the other hand, our proposal for building an emotion recognition model via keyboard typing stands out from others in the literature. First, the user is not asked to describe his current emotion using self-labeling techniques, which could be distracting in real-life scenarios and users may not identify their true emotion. However, this step is ensured by exposing very well studied elicitation videos along with accuracy questions. Second, the user's performance is taken into consideration. This was apprehended by converting the data into user-specific standardized values.

On a very first step, we started by conducting a Systematic Literature Review: it systematically summarizes key information in the area of emotion recognition through Keystroke dynamics and reports researchers' ongoing efforts towards creating a reference dataset. This could help researchers and practitioners to gain insight of the field and may help them find new lines of work.

Second, and due to the small sizes, scarcity, and non on-line availability of datasets, we developed and hosted a web application in order to collect data and build the dataset. Our dataset is so far the largest in terms of number of participants, number of instances, and is available online.

Third, the dataset is preprocessed and models are trained and tested.

The measures reflecting the recognition accuracy are exceptional compared to previous works. For instance, training the Res-Dataset using Random Committee algorithm, the True Positive Rate (Recall) reaches 95.7 % with F-score 95.1 % for "Calm" emotion. As for the accuracy rate and Kappa statistic, they reach 93.992 % and 0.9197, respectively, as the best recognition rates in this study.

One of the main limitations of this study is that we limited our scope to physical keyboards, although nowadays people are increasingly migrating to virtual keyboards. Moreover, the number of emotions used is low compared to what a computer user may experience. It is also important to admit that emotions may be expressed in various levels (e.g: angry, very angry, extremely angry). In addition, the texts used in the web application for the data collection process were written in English, reducing the scope of our targeted participants.

To address the above issues, many perspectives has arisen:

First, we can extend the approach to virtual keyboard and mobile devices. Second, we can explore the impact of some demographic effects (like gender and age) or typist types (one-hand or two-hand typing, and number of fingers involved) on the generated models.Third, an adaptive and incremental learning methodology will be developed, based on the proposed one, so when the trained model starts deteriorating, it can be retrained to update it to the new way of typing of the user. Finally, we can build an emotion recognition model that takes into consideration timing and frequency features along with previously discussed parameters such as gender and age. This model can be implemented in real-life scenarios like e-learning platforms. In closing, there are definitely many other possibilities for future work in this rapidly-developing field.

References

[1] R. W. Picard, Affective computing, MIT Press, Cambridge, Mass, 1997.

[2] J. S. Lee, Q. Xie, K. Lee, Informal digital learning of English and L2 willingness to communicate: roles of emotions, gender, and educational stage, Journal of Multilingual and Multicultural Development (2021) 1–17.

[3] S. Brave, C. Nass, Emotion in Human–Computer Interaction, The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications (2002).

[4] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, N. Arunkumar, Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS), IEEE Access 7 (2019) 57–67.

[5] J. Liu, H. Meng, M. Li, F. Zhang, R. Qin, A. K. Nandi, Emotion detection from EEG recordings based on supervised and unsupervised dimension reduction: Emotion Detection from EEG Recordings Based on Supervised and Unsupervised Dimension Reduction, Concurrency and Computation: Practice and Experience 30 (2018) e4446.

[6] D. J. Hemanth, J. Anitha, L. H. Son, Brain signal based human emotion analysis by circular back propagation and Deep Kohonen Neural Networks, Computers & Electrical Engineering 68 (2018) 170–180.

[7] S. Wioleta, Using physiological signals for emotion recognition, in: 2013 6th International Conference on Human System Interactions (HSI), IEEE, Sopot, Poland, 2013, pp. 556–561.

[8] M. Choraś, P. Mroczkowski, Keystroke Dynamics for Biometrics Identification, in: B. Beliczynski, A. Dzielinski, M. Iwanowski, B. Ribeiro (Eds.), Adaptive and Natural Computing Algorithms, volume 4432, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 424–431.

[9] R. Plutchik, The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice, American Scientist 89 (2001) 344–350.

[10] K. R. Scherer, What are emotions? And how can they be measured?, Social Science Information 44 (2005) 695–729.

[11] Y. Ma, Y. Hao, M. Chen, J. Chen, P. Lu, A. Košir, Audio-visual emotion fusion (AVEF): A deep efficient weighted approach, Information Fusion 46 (2019) 184–192.

[12] M. Kopaczka, J. Schock, J. Nestler, K. Kielholz, D. Merhof, A combined modular system for face detection, head pose estimation, face tracking and emotion recognition in thermal infrared images, in: 2018 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, Krakow, Poland, 2018, pp. 1–6.

[13] S. Zaghbani, N. Boujneh, M. S. Bouhlel, Facial Emotion Recognition for Adaptive Interfaces Using Wrinkles and Wavelet Network, in: 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), IEEE, Hammamet, 2017, pp. 342–349.

[14] Y.-Y. Ou, A.-C. Tsai, J.-F. Wang, P.-C. Lin, An integrated vision system on emotion understanding and identity confirmation, in: 2017 International Conference on Orange Technologies (ICOT), IEEE, Singapore, Singapore, 2017, pp. 205–209.

[15] T. Kundu, C. Saravanan, Advancements and recent trends in emotion recognition using facial image analysis and machine learning models, in: 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), IEEE, Mysuru, 2017, pp. 1–6.

[16] Q. Li, R. Gravina, G. Fortino, Posture and Gesture Analysis Supporting Emotional Activity Recognition, in: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Miyazaki, Japan, 2018, pp. 2742–2747.

[17] J. Kim, J. Seo, T. H. Laine, Detecting boredom from eye gaze and EEG, Biomedical Signal Processing and Control 46 (2018) 302–313.

[18] D. B. Setyohadi, S. Kusrohmaniah, S. B. Gunawan, P. Pranowo, A. S. Prabuwono, Galvanic Skin Response Data Classification for Emotion Detection, International Journal of Electrical and Computer Engineering (IJECE) 8 (2018) 4004.

[19] B. Ramayah, Detecting Emotional Changes among Visually Impaired Users in Complex and Non-Complex Layout using Neuro-Physiological Signal, in: Proceedings of the 2018 7th International Conference on Software and Computer Applications - ICSCA 2018, ACM Press, Kuantan, Malaysia, 2018, pp. 302–305.

[20] D. Girardi, F. Lanubile, N. Novielli, Emotion detection using noninvasive low cost sensors, in: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, San Antonio, TX, 2017, pp. 125–130.

[21] M. Soleymani, M. Mortillaro, Behavioral and Physiological Responses to Visual Interest and Appraisals: Multimodal Analysis and Automatic Recognition, Frontiers in ICT 5 (2018) 17.

[22] K. Sailunaz, M. Dhaliwal, J. Rokne, R. Alhajj, Emotion detection from text and speech: a survey, Social Network Analysis and Mining 8 (2018) 28.

[23] R. Lotfian, C. Busso, Lexical Dependent Emotion Detection Using Synthetic Speech Reference, IEEE Access 7 (2019) 22071–22085.

[24] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, A. E. Hassanien, Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media, Applied Soft Computing 97 (2020) 106754.

[25] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, P. Agrawal, Understanding Emotions in Text Using Deep Learning and Big Data, Computers in Human Behavior 93 (2019) 309–317.

[26] S. Salmeron-Majadas, R. S. Baker, O. C. Santos, J. G. Boticario, A Machine Learning Approach to Leverage Individual Keyboard and Mouse Interaction Behavior From Multiple Users in Real-World Learning Scenarios, IEEE Access 6 (2018) 39154–39179.

[27] J. Kim, H. Kim, P. Kang, Keystroke dynamics-based user authentication using freely typed text based on user-adaptive feature extraction and novelty detection, Applied Soft Computing 62 (2018) 1077–1087.

[28] A. Pentel, Emotions and user interactions with keyboard and mouse, in: 2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA), IEEE, Larnaca, 2017, pp. 1–6.

[29] I. Tsimperidis, A. Arampatzis, A. Karakos, Keystroke dynamics features for gender recognition, Digital Investigation 24 (2018) 4–10.

[30] A. Daribay, M. S. Obaidat, P. V. Krishna, Analysis of Authentication System Based on Keystroke Dynamics, in: 2019 International Conference on Computer, Information and Telecommunication Systems (CITS), IEEE, Beijing, China, 2019, pp. 1–6.

[31] Y. Zhong, Y. Deng, A Survey on Keystroke Dynamics Biometrics: Approaches, Advances, and Evaluations, in: Y. Zhong, Y. Deng (Eds.), Gate to Computer Science and Research, volume 2, 1st ed., Science Gate Publishing P.C., 2015, pp. 1–22.

[32] P. Ekman, An argument for basic emotions, Cognition and Emotion 6 (1992) 169–200.

[33] C. E. Izard, Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm, Perspectives on Psychological Science 2 (2007) 260–280.

[34] J. A. Russell, Core affect and the psychological construction of emotion., Psychological Review 110 (2003) 145–172.

[35] I. Bakker, T. van der Voordt, P. Vink, J. de Boon, Pleasure, Arousal, Dominance: Mehrabian and Russell revisited, Current Psychology 33 (2014) 405–421.

[36] J. R. Fontaine, K. R. Scherer, E. B. Roesch, P. C. Ellsworth, The World of Emotions is not Two-Dimensional, Psychological Science 18 (2007) 1050–1057.

[37] B. N. Colby, A. Ortony, G. L. Clore, A. Collins, The Cognitive Structure of Emotions., Contemporary Sociology 18 (1989) 957.

[38] M. G. Valdez, J.-J. Merelo, A. H. Aguila, A. M. Soto, Mining of Keystroke and Mouse Dynamics to Increase the Engagement of Students with Programming Assignments, in: C. Sabourin, J. J. Merelo, K. Madani, K. Warwick (Eds.), Computational Intelligence, volume 829, Springer International Publishing, Cham, 2019, pp. 41–61.

[39] A. Maalej, I. Kallel, Does Keystroke Dynamics tell us about Emotions? A Systematic Literature Review and Dataset Construction, in: 2020 16th International Conference on Intelligent Environments (IE), IEEE, Madrid, Spain, 2020, pp. 60–67.

[40] K. Bakhtiyari, M. Taghavi, H. Husain, Hybrid affective computing—keyboard, mouse and touch screen: from review to experiment, Neural Computing and Applications 26 (2015) 1277–1296.

[41] C. Epp, M. Lippold, R. L. Mandryk, Identifying emotional states using keystroke dynamics, in: Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11, ACM Press, Vancouver, BC, Canada, 2011, p. 715.

[42] A. Kolakowska, Recognizing emotions on the basis of keystroke dynamics, in: 2015 8th International Conference on Human System Interaction (HSI), IEEE, Warsaw, Poland, 2015, pp. 291–297.

[43] A. N. H. Nahin, J. M. Alam, H. Mahmud, K. Hasan, Identifying emotion by keystroke dynamics and text pattern analysis, Behaviour & Information Technology 33 (2014) 987–996.

[44] L. M. Vizer, Different strokes for different folks: individual stress response as manifested in typed text, in: CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13, ACM Press, Paris, France, 2013, p. 2773.

[45] R. Shikder, S. Rahaman, F. Afroze, A. B. M. A. A. Islam, Keystroke/mouse usage based emotion detection and user identification, in: 2017 International Conference on Networking, Systems and Security (NSysS), IEEE, Dhaka, Bangladesh, 2017, pp. 96–104.

[46] P.-M. Lee, W.-H. Tsui, T.-C. Hsiao, The influence of emotion on keyboard typing: an experimental study using visual stimuli, BioMedical Engineering OnLine 13 (2014) 81.

[47] Emotion Recognition through Keystroke Dynamics on Touchscreen Keyboards:, in: Proceedings of the 15th International Conference on Enterprise Information Systems, SciTePress - Science and and Technology Publications, Angers, France, 2013, pp. 31–37.

[48] J. A. Harris, D. Isaacowitz, Emotion in Cognition, in: International Encyclopedia of the Social & Behavioral Sciences, Elsevier, 2015, pp. 461–466.

[49] T. L. Gilman, R. Shaheen, K. M. Nylocks, D. Halachoff, J. Chapman, J. J. Flynn, L. M. Matt, K. G. Coifman, A film set for the elicitation of emotion in research: A comprehensive catalog derived from four decades of investigation, Behavior Research Methods 49 (2017) 2061–2082.

[50] L. Fernández-Aguilar, B. Navarro-Bravo, J. Ricarte, L. Ros, J. M. Latorre, How effective are films in inducing positive and negative emotional states? A meta-analysis, PLOS ONE 14 (2019) e0225040.

[51] R. A. Stevenson, J. A. Mikels, T. W. James, Characterization of the Affective Norms for English Words by discrete emotional categories, Behavior Research Methods 39 (2007) 1020–1024.

[52] I. H. Witten, E. Frank, Data Mining Practical Machine Learning Tools and Techniques, Elsevier Science & Technology Books, San Diego, CA, USA, 0000. OCLC: 892785243.

[53] I. F. Ilyas, X. Chu, Data cleaning, number 28 in ACM Books, first edition ed., Association for Computing Machinery, New York, NY, 2019.