

Using Mouse Dynamics to Assess Stress During Online Exams

Davide Carneiro¹(✉), Paulo Novais¹, José Miguel Pêgo^{2,3}, Nuno Sousa^{2,3},
and José Neves¹

¹ Algorimti Centre, University of Minho, Braga, Portugal
{dcarneiro,pjon,jneves}@di.uminho.pt

² School of Health Sciences, Life and Health Sciences Research Institute (ICVS),
University of Minho, Braga, Portugal

{jmpego,njcsousa}@ecsaude.uminho.pt

³ ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal

Abstract. Stress is a highly complex, subjective and multidimensional phenomenon. Nonetheless, it is also one of our strongest driving forces, pushing us forward and preparing our body and mind to tackle the daily challenges, independently of their nature. The duality of the effects of stress, that can have positive or negative effects, calls for approaches that can take the best out of this biological mechanism, providing means for people to cope effectively with stress. In this paper we propose an approach, based on mouse dynamics, to assess the level of stress of students during online exams. Results show that mouse dynamics change in a consistent manner as stress settles in, allowing for its estimation from the analysis of the mouse usage. This approach will allow to understand how each individual student is affected by stress, providing additional valuable information for educational institutions to efficiently adapt and improve their teaching processes.

Keywords: Stress · Human-computer interaction · e-Learning · Mouse dynamics

1 Introduction

Modern life can be often assume a frenetic rhythm, caused by competitiveness, social judgement, productivity demands, information overload and many other modern sources of pressure. This exerts a significant and constant pressure on individuals, driving them to a constant attempt to perform more and better. There are environments which constitute particularly “good” examples of this reality. The workplace, for one, is a milieu currently associated to stress, competition, demanding working conditions and even certain illnesses. The classroom is another one in which individuals, from early in their lives, are confronted with frequent evaluations of their performance and the pressure that stems from its impact on their future and from the social judgement of their peers [3].

Higher education, in particular, is a period of the individual's educational path that is especially prone to result in added pressure [4]. It is so because it constitutes a transition period before students reach the working environment, combining the fears and the pressure of both environments. Students are subjected to increasing periods of work with a progressive focus on autonomy and continuous assessment as mandated by current educational policies. The increasing workload is perceived as stressful and commonly leads to mental disorders and perception that their cognitive performance is below their expected standards [5]. This is corroborated by the high prevalence of anxiety disorders among higher education students.

Assessment is a fundamental phase in the training and certification process that a higher education student is submitted to. It is also one of the strongest stress factors due to the high-stake implications in the academic progress and self-perceived image. Stress is a risk factor for anxiety and may lead to worsening of performance in assessment tasks [7, 13].

This paper focuses on this kind of specific moments, which may be extremely important in the individuals' lives and, therefore, extremely stressful. Specifically, the paper discusses how a group of medical students was monitored in order to study the effect of stress/anxiety in the performance of high demand tasks. Students were monitored in terms of the efficiency of their interaction patterns with the computer, an approach that can be included in the so-called behavioural biometrics.

Results show that, in a general way, the performance of the interaction increases with stress. However, the study carried out also points out that not all students behave alike and that individual behavioural models should be developed for increased accuracy.

The main aim of this line of research is to provide additional sources of contextual information about students during evaluation tasks that can allow the educational institution to design better and more individualized teaching strategies.

The remaining of the paper is organized as follows. Section 2 briefly presents some related work on the field of Behavioral Biometrics. Section 3 details the experimental study carried out, followed by a statistical analysis of the data in Sect. 4. Section 5 details the training of classifiers for stress assessment and Sect. 5 is dedicated to the discussion of the results and the future work.

2 Related Work

Biometrics consist of the use of individual's characteristics (usually physical or physiological), generally for the purpose of user identification or access control. These characteristics include fingerprints, facial recognition, retina or iris recognition or even DNA analysis [8]. More recently research started in the so-called behavioural biometrics, which is based on behavioural traits of the individual, including interaction or movement patterns, speech rhythm or movement patterns.

A general and thorough review of behavioural biometrics, addressing several fields of applications, was conducted by Yampolskiy et al. [9]. One of the most common fields of application is for security-related purposes. Shen et al. detail a system for identifying and authenticating the user of a computer based on his usage of the mouse [10]. Other similar systems are detailed in [8]. Our own research team has developed work in the past to assess attention and fatigue from behavioural biometrics [1].

Specifically, the use of the mouse, known as Mouse Dynamics, has been used with success in numerous approaches [11, 12] and has shown that it can produce a wide range of different features. In this paper we also make use of Mouse Dynamics, proposing a set of 10 different features to characterize the interaction of medical students with the computer while taking exams.

3 Study Design

The purpose of this work is to determine if increasing levels of stress on medical students have a significant effect on mouse dynamics while they participate in high stake exams using a computer. For this purpose, a group of 53 students was selected. The participation in the study did not imply any change in their routines, i.e., these students needed to perform exactly the same tasks as those not taking part in the study.

In this kind of exams, when students enter the room, they are indicated their seats. At the designated time they log in the exam platform using their personal credentials and the exam begins. During the exam, which consists mostly of single-best-answer multiple choice questions [6], students use mostly the mouse as an interaction means. When the exam ends, students are allowed to leave the room.

The collection of the interaction data is completely transparent from the point of view of the student. It is performed by a previously developed application described in [1] which runs in the background, capturing all system events related to interaction. Subsections 3.1 and 3.2 describe, respectively, the process of data collection and its transformation into useful behavioural features.

3.1 Data Collection

The data collection tool, which is installed locally in each computer, runs in the background and listens to all the system events related to the use of the mouse. As the events happen it continuously builds a log that is sent to a centralized server, allowing a posterior analysis of the data. The log includes the following events and the respective information:

- MOV, timestamp, posX, posY
An event describing the movement of the mouse, in a given time, to coordinates (posX, posY) in the screen;

- `MOUSE_DOWN`, timestamp, [Left|Right], posX, posY
This event describes the first half of a click (when the mouse button is pressed down), in a given time. It also describes which of the buttons was pressed (left or right) and the position of the mouse in that instant;
- `MOUSE_UP`, timestamp, [Left|Right], posX, posY
An event similar to the previous one but describing the second part of the click, when the mouse button is released;
- `MOUSE_WHEEL`, timestamp, dif
This event describes a mouse wheel scroll of amount dif, in a given time;
- `KEY_DOWN`, timestamp, key
Identifies a given key from the keyboard being pressed down, at a given time;
- `KEY_UP`, timestamp, key
Describes the release of a given key from the keyboard, in a given time.

The following example depicts a brief log that starts with some mouse movement (first two lines), contains a click with a little drag (lines 3–5) and ends with some more movement (last two lines).

```
MOV, 635296941683402953, 451, 195
MOV, 635296941684123025, 451, 197
MOUSEDOWN, 635296941684443057, Left, 451, 199
MOV, 635296941685273140, 452, 200
MOUSE_UP, 635296941685283141, Left, 452, 200
MOV, 635296941685723185, 452, 203
MOV, 635296941685803193, 454, 205
```

3.2 Extraction of Behavioural Features

The individual logs build by the aforementioned application are then processed in order to compile information that can efficiently characterize the behaviour of students while interacting with the computer. This subsection details the features extracted from the logs of the students detailed in Sect. 3.2.

It is important to note that these features aim at quantifying the students' performance. Taking as example the movement of the mouse, one never moves it in a straight line between two points, there is always some degree of curve. The larger the curve, the less efficient the movement is. An interesting property of these features is that, except for mouse velocity and acceleration, an increasing value denotes a decreasing performance (e.g. longer click \Rightarrow poorer performance, larger average excess of distance \Rightarrow poorer performance). Concerning mouse velocity and acceleration, the relationship is not straightforward. While up to a certain point they might indicate better performance, after that point people have a smaller degree of control, i.e., less precision. For that reason, and given that the focus of this work is on assessing performance, these two features will not be considered in the data analysis.

The following features are considered:

Absolute Sum of Angles (ASA)

UNITS - degrees

This feature seeks to find how much the mouse “turned”, independently of the direction to which it turned (Fig. 1(a)). In that sense, it is computed as the absolute of the value returned by function $degree(x1, y1, x2, y2, x3, y3)$, as depicted in Eq. 1.

$$rCls_angle = \sum_{i=0}^{n-2} | degree(posx_i, posy_i, posx_{i+1}, posy_{i+1}, posx_{i+2}, posy_{i+2}) | \tag{1}$$

Average Distance of the Mouse to the Straight Line (ADMSL)

UNITS - pixels

This feature measures the average distance of the mouse to the straight line defined between two consecutive clicks. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, mup and mdo , respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors $posx$ and $posy$, of size n , holding the coordinates of the consecutive MOUSE_MOV events between mup and mdo . The sum of the distances between each position and the straight line defined by the points $(x1, y1)$ and $(x2, y2)$ is given by Eq. 2, in which $ptLineDist$ returns the distance between the specified point and the closest point on the infinitely-extended line defined by $(x1, y1)$ and $(x2, y2)$. The average distance of the mouse to the straight (Fig. 1(b)) line defined by two consecutive clicks is this given by s_dists/n .

$$s_dists = \sum_{i=0}^{n-1} ptLineDist(posx_i, posy_i) \tag{2}$$

Average Excess of Distance (AED)

UNITS - pixels

This feature measures the average excess of distance that the mouse travelled between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let

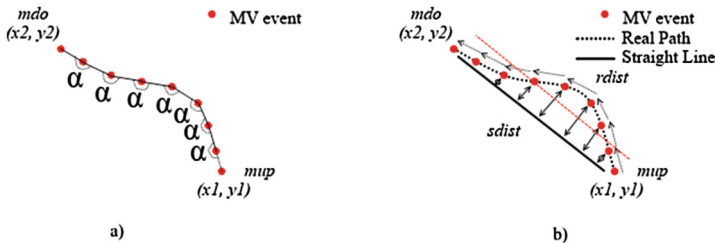


Fig. 1. (a) The sum of the angles of the mouse’s movement is given by summing all the angles between each two consecutive movement vectors. (b) The average distance at which the mouse is from the shortest line between two clicks is depicted by the straight dashed line.

us assume two consecutive MOUSE_UP and MOUSE_DOWN events, mup and mdo , respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. To compute this feature, first it is measured the distance in straight line between the coordinates of mup and mdo as $s_dist = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$. Then, it is measured the distance actually travelled by the mouse by summing the distance between each two consecutive MOUSE_MV events. Let us assume two vectors $posx$ and $posy$, of size n , holding the coordinates of the consecutive MOUSE_MV events between mup and mdo . The distance actually travelled by the mouse, $real_dist$ is given by Eq. 3. The average excess of distance between the two consecutive clicks (Fig. 2(a)) is thus given by r_dist/s_dist .

Click Duration (CD)

UNITS - milliseconds

Measures the timespan between two consecutive MOUSE_UP and MOUSE_DOWN events.

Distance Between Clicks (DBC)

UNITS - pixels

Represents the total distance travelled by the mouse between two consecutive clicks, i.e., between each two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, mup and mdo , respectively in the coordinates $(x1, y1)$ and $(x2, y2)$. Let us also assume two vectors $posx$ and $posy$, of size n , holding the coordinates of the consecutive MOUSE_MOV events between mup and mdo . The total distance travelled by the mouse is given by Eq. 3.

$$r_dist = \sum_{i=0}^{n-1} \sqrt{(posx_{i+1} - posx_i)^2 + (posy_{i+1} - posy_i)^2} \quad (3)$$

Distance of the Mouse to the Straight Line (DMSL)

UNITS - pixels

This feature is similar to the previous one in the sense that it will compute the s_dists between two consecutive MOUSE_UP and MOUSE_DOWN events, mup and mdo , according to Eq. 2. However, it returns this sum rather than the average value during the path.

Excess of Distance (ED)

UNITS - pixels

This feature measures the excess of distance that the mouse travelled between each two consecutive MOUSE_UP and MOUSE_DOWN events. r_dist and s_dist are computed as for the AED feature. However, ED is given by $r_dist - s_dist$

Mouse Acceleration (MA)

UNITS - pixels/milliseconds²

The velocity of the mouse (in pixels/milliseconds) over the time (in milliseconds). A value of acceleration is computed for each interval defined by two consecutive

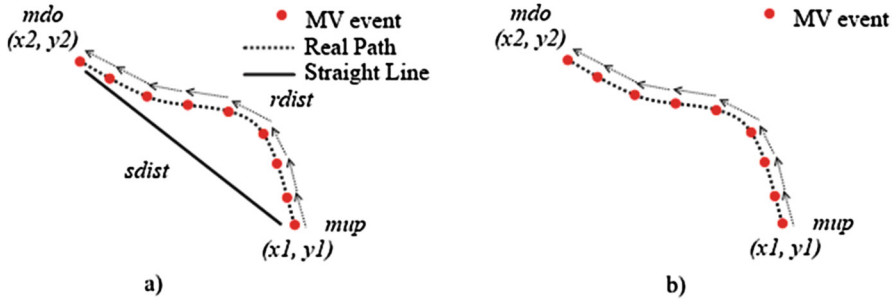


Fig. 2. (a) A series of MOV events, between two consecutive clicks of the mouse. The difference between the shortest distance (*sdist*) and distance actually travelled by the mouse (*rdist*) is depicted. (b) The real distance travelled by the mouse between each two consecutive clicks is given by summing the distances between each two consecutive MOV events.

MOUSE_UP and MOUSE_DOWN events, using the intervals and data computed for the Velocity.

Mouse Velocity (MV)

UNITS - pixels/milliseconds

The distance travelled by the mouse (in pixels) over the time (in milliseconds). The velocity is computed for each interval defined by two consecutive MOUSE_UP and MOUSE_DOWN events. Let us assume two consecutive MOUSE_UP and MOUSE_DOWN events, *mup* and *mdu*, respectively in the coordinates (x_1, y_1) and (x_2, y_2) , that took place respectively in the instants $time_1$ and $time_2$. Let us also assume two vectors *posx* and *posy*, of size n , holding the coordinates of the consecutive MOUSE_MOVE events between *mup* and *mdu*. The velocity between the two clicks is given by $r_dist / (time_2 - time_1)$, in which *r_dist* represents the distance travelled by the mouse and is given by Eq. 3.

Time Between Clicks (TBC)

UNITS - milliseconds

The timespan between two consecutive MOUSE_UP and MOUSE_DOWN events, i.e., how long did it took the individual to perform another click.

4 Data Analysis

After the collection and transformation of the data, work continued to the analysis of the collected data. The first phase of this analysis, in which statistical methods are used to obtain preliminary conclusions, is described in this section. The second phase, in which machine learning methods are used to model the students' response to stress, is described in Sect. 5.

Data was analysed in two different ways. First, a general analysis was carried out with the aim of searching for group trends, i.e., behaviours common to

a significantly large slice of the population. Secondly, an individual analysis was performed, aiming at the analysis of each student at a time, in order to understand their differences and support the development of personalized models.

In both analysis the features Mouse Velocity and Mouse Acceleration were not considered, for the reasons pointed out in Sect. 3.2.

Group performance was studied conducting a feature-by-feature analysis of the data, for all the participants. For each one of the 53 participants it was computed the correlation of each feature with time, for the whole duration of the exam. Figure 3 shows the distribution of the values of correlation for all the users and for each feature. It can be inferred that correlations are mostly negative (with the exception of mouse velocity and acceleration). This provides a first hint towards the trend that most of the students increase their performance with stress.

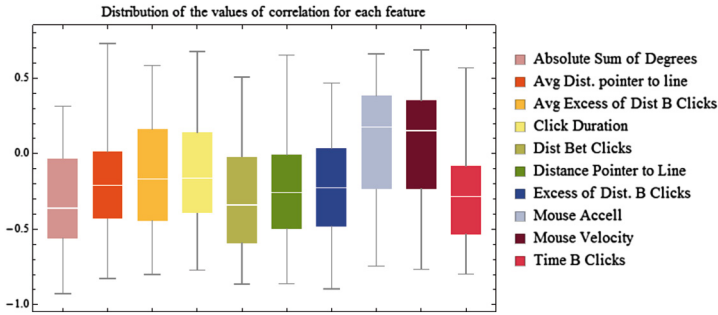


Fig. 3. Distribution of the participant’s values of correlation with time, for each feature.

Indeed, most of the participants evidence this negative correlation for most of the features. Figure 4 shows the values of the features during the exam, for three arbitrary participants. It can be seen that, despite variations in the values of correlation, they are mostly negative.

Having established that performance tends to increase with time, the next step was to partition the data in intervals, so as to compare, for instance, the distributions of the first and the last interval in order to determine if there are statistically significant differences between them. Five intervals were defined, of equal length. Data from all features were divided accordingly. From this we concluded that the average value in each part tends to decrease with time for all the features (Table 1).

The last column of Table 1 shows the percentage of participants, for each feature, for whom the differences of the data between Part 1 and Part 5 are statistically significant. For this purpose, the Mann-Whitney test was used, with a *p*-value of 0.05.

The group analysis performed showed that a certain trend of improving performance for the group of participants can be found in the data. However, the

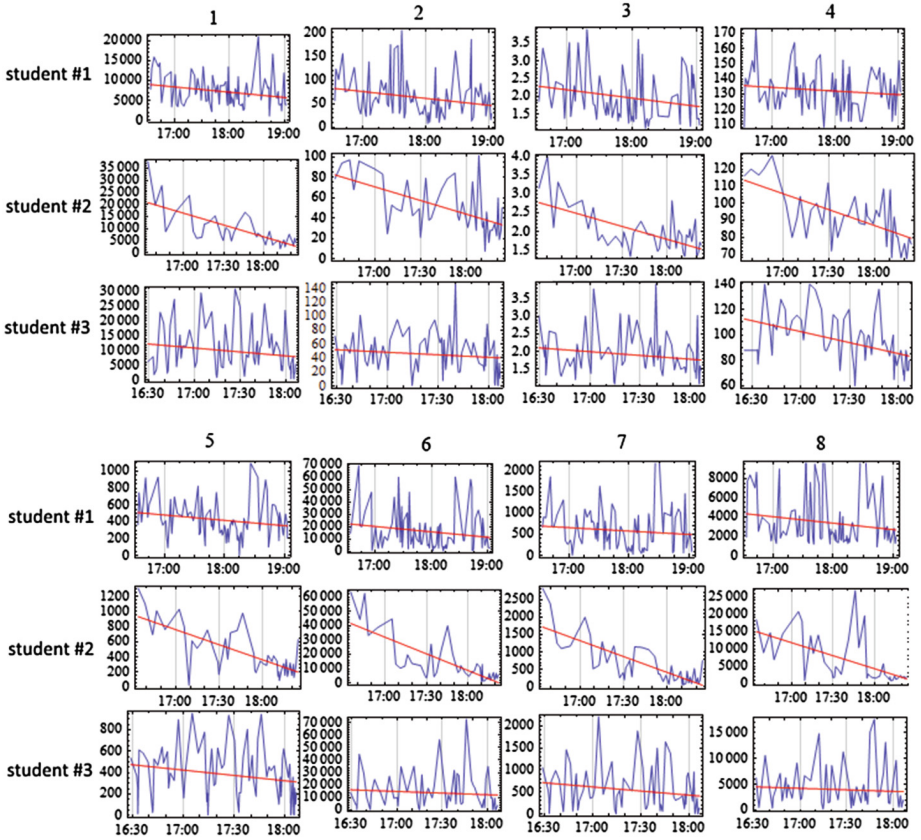


Fig. 4. Time plot of the features for three arbitrary students. The negative correlation with time is visible for all features. Lines depict three different students. Columns depict the following eight features: 1 - ASA, 2 - ADMSL, 3 - AED, 4 - CD, 5 - DBC, 6 - DMSL, 7 - ED, 8 - TBC.

collected data also shows that not all students behave alike: there are cases in which correlation is stronger/weaker and there are cases in which correlation is positive for some features.

Indeed, as the results presented further ahead show, individually trained classifiers perform better than a general one. Figure 5 details time plots for a specific student. The negative correlation in all the features is clearly visible. The values of correlation for the plots in the Figure are, respectively, -0.926177 , -0.819893 , -0.789006 , -0.534707 , -0.863852 , -0.861359 , -0.895728 and -0.685151 .

5 Classifying Stress from the Interaction with the Mouse

The data collected and already detailed in the previous sections was used to train two classifiers. The first was trained with data from all the participants.

Table 1. Average value for each part of the data and percentage of participants for which there is a statistically significant difference between the distributions of Part 1 and Part 5 (Mann-Whitney, p -value < 0.05).

Feature	Part 1	Part 2	Part 3	Part 4	Part 5	% MannWhit 1-5
ASA	11689.00	11052.30	10496.60	9822.63	8488.60	54.72 %
ADMSL	81.93	76.02	69.13	68.25	69.67	49.06 %
AED	2.99	2.92	2.83	2.73	2.82	35.85 %
CD	126.74	125.85	124.21	120.92	119.51	47.17 %
DBC	600.76	557.72	499.30	442.34	403.83	66.04 %
DMSL	32812.60	31001.50	29811.20	25013.90	24667.60	45.28 %
ED	1143.04	1052.62	924.77	871.06	831.93	49.06 %
TBC	10581.80	10079.50	9674.05	8896.50	7526.92	50.94 %

The second was trained using data from a single user, in an attempt to compare the general model with a specific one in terms of performance. The process followed for constructing the dataset was similar for both cases. For each feature we selected the first and the last of the five groups of data. Instances from the first group were labelled as 'calm' whereas instances from the second group were labelled as 'stress'.

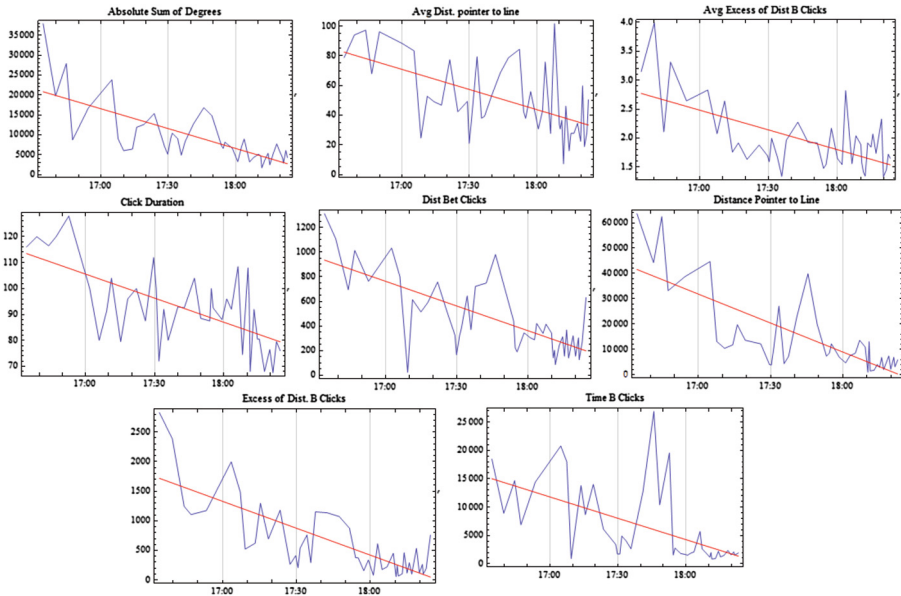


Fig. 5. Plot of the data of all features of a particular student. Negative correlations are visible. Features, from bottom to top and left to right: 1 - ASA, 2 - ADMSL, 3 - AED, 4 - CD, 5 - DBC, 6 - DMSL, 7 - ED, 8 - TBC.

The first dataset contains a total of 2438 instances whereas the second one contains a total of 162 instances. In both cases a Naive Bayes Classifier was used, with 10-fold cross-validation [2]. This is a classification algorithm that resulted from the combination of Decision Trees and Naive Bayes based algorithms. It generally outperforms these algorithms alone and tends to scale better to large databases.

The model trained with the general dataset resulted in a tree of size 13, with 7 leaves. The number of correctly classified instances amounts to 1575 (64.6%), which is not an especially satisfying value. We must however recall that this includes data from all the students, integrating and blurring all their individual differences.

Indeed, individually trained classifiers tend to have much more satisfying results. Taking as example the case depicted in Fig. 5, the percentage of correctly classified instances rises to 86.4% (140 out of 162 instances).

6 Discussion and Future Work

This work allows drawing some interesting conclusions about students and their behaviour during exams. First of all, it shows that stress actually influences students' interaction patterns with the computer during high-stake exams. This could open the door to the development of new approaches for assessing and managing stress, especially in the context of superior education. It also shows that it is possible to train classifiers that can carry out this task in real-time. Moreover, it is our conviction that this kind of approaches can be extended to other domains, namely the workplace. Nonetheless, this calls for the carrying out of new studies since the characteristics of these milieus are imminently different.

There are, however, many issues still to address. The first is to understand why some students improve their general performance with stress while others do not. Knowing which factors influence this might allow higher education institutions to implement individualized coping strategies with the aim to mitigate negative stress effects. This will be addressed in future work, by comparing the extensive profile of students maintained by the School of Health Sciences with the performance features.

We will also improve the training of the classifiers by selecting the most significant features only. Moreover, using clustering techniques, we will try to identify groups of students that behave alike. The main advantages that we expect from this are twofold. (1) We will be able to train group classifiers that have performances similar to the ones trained for individual students. (2) Students who are participating for the first time, for whom there is still not a model trained, can be assigned to a known group with a similar behavioural pattern, thus using the model of that group for classifying his own behaviour.

Acknowledgments. This work is part-funded by ERDF - European Regional Development Fund through the COMPETE Programme (operational programme

for competitiveness) and by National Funds through the FCT (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-028980 (PTDC/EEI-SII/1386/2012) and project PEst-OE/EEI/UI0752/2014.

References

1. Pimenta, A., Carneiro, D., Novais, P., Neves, J.: Detection of distraction and fatigue in groups through the analysis of interaction patterns with computers. In: Camacho, D., Braubach, L., Ventcinque, S., Badica, C. (eds.) *Intelligent Distributed Computing VIII*. SCI, vol. 570, pp. 29–40. Springer, Heidelberg (2014)
2. Kohavi, R.: Scaling up the accuracy of naive-bayes classifiers: a decision-tree hybrid. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (1996)
3. Dyrbye, L.N., Thomas, M.R., Shanafelt, T.D.: Systematic review of depression, anxiety, and other indicators of psychological distress among US and Canadian medical students. *Acad. Med.* **81**(4), 354–373 (2006)
4. Goebert, D., Thompson, D., Takeshita, J., Beach, C., Bryson, P., Ephgrave, K., Kent, A., Kunkel, M., Schechter, J., Tate, J.: Depressive symptoms in medical students and residents: a multischool study. *Acad. Med.* **84**(2), 236–241 (2009)
5. Strous, R.D., Shoenfeld, N., Lehman, A., Wolf, A., Snyder, L., Barzilai, O.: Medical students' self-report of mental health conditions. *Int. J. Med. Educ.* **3**, 1 (2012)
6. Case, S.M., Swanson, D.B.: *Constructing Written Test Questions for the Basic and Clinical Sciences*. National Board of Medical Examiners, Philadelphia (1998)
7. McEwen, B.S.: Brain on stress: how the social environment gets under the skin. *Proc. Natl. Acad. Sci.* **109**, 17180–17185 (2012)
8. Jain, A.K., Ross, A., Prabhakar, S.: An introduction to biometric recognition. *IEEE Trans. Circuits Syst. Video Technol.* **14**(1), 4–20 (2004)
9. Yampolskiy, R.V., Govindaraju, V.: Behavioral biometric: a survey and classification. *Int. J. Biometrics* **1**(1), 81–113 (2008)
10. Shen, C., Cai, Z., Guan, X., Du, Y., Maxion, R.A.: User authentication through mouse dynamics. *IEEE - Inf. Forensics Secur.* **8**(1), 16–30 (2013)
11. Ahmedand, A.A.E., Traore, I.: Anomaly intrusion detection based on biometrics. In: *Proceedings of IEEE Information Assurance Workshop*, pp. 452–453, West Point (2005)
12. Aksari, Y., Artuner, H.: Active authentication by mouse movements. In: *Proceedings of 24th International Symposium on Computer and Information Science*, pp. 571–574, Guzelyurt (2009)
13. Soares, J.M., Sampaio, A., Ferreira, L.M., Santos, N.C., Marques, F., Palha, J.A., Cerqueira, J.J., Sousa, N.: Stress-induced changes in human decision-making are reversible. *Transl. psychiatry* **2**(7), e131 (2012)