Mouse dynamics correlates to student behaviour in computer-based exams

DAVIDE CARNEIRO*, *CIICESI, ESTG,Polytechnic Institute of Porto, Felgueiras, Portugal; Algorimti Centre, University of Minho, Braga, Portugal.*

PAULO NOVAIS**, *Algorimti Centre, University of Minho, Braga, Portugal.*

NUNO SOUSA†, *Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal; ICVS/3B's - PT Government Associate Laboratory Braga/Guimarães, Portugal.*

JOSÉ MIGUEL PÊGO‡, *Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal; ICVS/3B's - PT Government Associate Laboratory Braga/Guimarães, Portugal.*

JOSÉ NEVES§, *Algorimti Centre, University of Minho, Braga, Portugal.*

Abstract

Nowadays, it is common for higher education institutions to use computer-based exams, partly or integrally, in their evaluation processes. These exams, much like their paper-based counterparts, are one of the most significant sources of stress in the life of students. However, the fact that exams are undertaken in a computer allows for new features to be acquired that may provide more reliable insights into the behaviour and state of the student during the exam. In this article we analyse these novel behavioural features and explore, to which extent, they can point out previously unknown phenomena. Specifically, we show that the time a student takes to complete an exam is correlated with mouse dynamics features. In practical terms, we are able to predict the duration of each individual exam with a satisfying error based on the interaction patterns of the student.

Keywords: Human–computer interaction, stress, mouse dynamics, student behaviour.

1 Introduction

Modern life can 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 with stress, competition, demanding work conditions and even certain illnesses. The classroom is another one in which individuals, from early

*E-mail: dcarneiro@di.uminho.pt
**E-mail: pjon@di.uminho.pt
†E-mail: jneves@di.uminho.pt
‡E-mail: jmpego@ecsaude.uminho.pt
§E-mail: njcsousa@ecsaude.uminho.pt

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 [5].

Higher education, in particular, is a period of the individual's educational path that is especially prone to result in added pressure [6]. 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 [13]. 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-stakes 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 [9, 12].

This article focuses on these kind of specific moments, which may be extremely important in the individuals' lives and, therefore, extremely stressful. Specifically, the article 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 the way the behaviour of the student changes during the exam is related to the duration of the exam. The potential consequences of this finding are twofold:

– The possibility of predicting the duration of each exam, in real-time, during the carrying out of the exam;
– The finding of new previously unknown variables that can potentially relate to psychological traits of the students, such as their ability to cope with stress.

The remaining of the article is organized as follows. Section 2 presents some related work, namely in the fields of Biometrics and Mouse Dynamics. Section 3 describes the experimental study carried out as well as the system that enabled the collection of the data and the interaction features that it supports. Section 4 describes how data were analysed. Section 5 details how the duration of the exam can be predicted from the student's interaction patterns and Section 6 discusses the findings and points out future steps.

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 [7]. More recently research started in the so-called behavioural biometrics, which is based on behavioural traits of the individual, including interaction or movement patterns and speech rhythm or movement patterns.

A general and thorough review of behavioural biometrics, addressing several fields of applications, was conducted by Yampolskiy and Govindaraju [14]. One of the most common fields of application is for security-related purposes. Shen et al. [11] detail a system for identifying and authenticating the user of a computer based on his usage of the mouse. Other similar systems are detailed in [7]. Our own research team has developed work in the past to assess attention and fatigue from behavioural biometrics [10].

Specifically, the use of the mouse, known as Mouse Dynamics, has been used with success in numerous approaches [1, 2] and has shown that it can produce a wide range of different features. In this article, 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 research is to study how the interaction patterns of students evolve during particularly stressful moments such as high-stakes computer-based exams.

For this purpose, a cohort of 104 students was selected. When students enter the room, they are indicated their seats. At the designated time they log in the exam platform (medQuizz®, iCognitus) using their personal credentials and the exam begins. During the exam, which consists mostly of single-best-answer multiple choice questions [4], students use almost exclusively the mouse as an interaction means. This is, indeed, a very specific type of exam, which is also common in Medicine. Moreover, the results obtained may not be replicated in exams with different characteristics, namely in exams that do not depend so significantly on the use of the mouse. Similar studies must thus be conducted in the future using exams with different characteristics.

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 [10] which runs in the background, capturing all system events related to interaction. This means that the students do not need to perform any specific interaction.

The participants took the same exam simultaneously, using similar computers. Nonetheless, they were allowed to leave the exam at any moment: either because they finished the exam or because they gave up.

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
MOUSE_DOWN, 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 built 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 Section 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
UNITS - degrees
Absolute Sum of Angles (ASA) seeks to find how much the mouse 'turned', independently of the direction to which it turned (Figure 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 (1).

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

AVERAGE DISTANCE OF THE MOUSE TO THE STRAIGHT LINE
UNITS - pixels
Average distance of the mouse to the straight line (ADMSL) 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 (2), in which
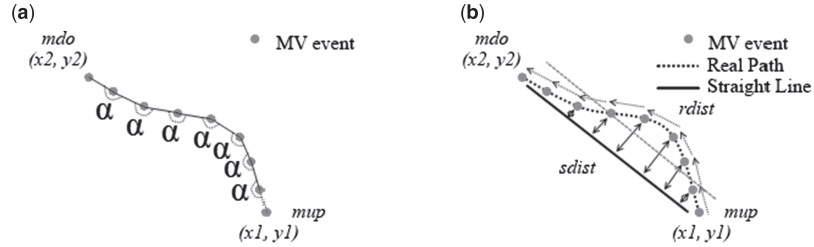
FIG. 1. (a) The sum of the angles of the mouse's movement is given by summing all the angles between each two consecutive movement vectors. (b) The average distance at which the mouse is from the shortest line between two clicks is depicted by the straight dashed line.

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 (Figure 1 (b)) line defined by two consecutive clicks is thus given by s_dists/n.

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

AVERAGE EXCESS OF DISTANCE
UNITS - pixels
Average excess of distance (AED) measures the average excess of distance that the mouse travelled between each two consecutive MOUSE_UP and MOUSE_DOWN events is measured. Let 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 the distance in straight line between the coordinates of *mup* and *mdo* as $s\_dist = \sqrt{(x2-x1)^2 + (y2-y1)^2}$ is measured. Then, the distance actually travelled by the mouse by summing the distance between each two consecutive MOUSE_MV events is measured. 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 equation (3). The average excess of distance between the two consecutive clicks (Figure 2 (a)) is thus given by r_dist/s_dist.

CLICK DURATION
UNITS - milliseconds
Click duration (CD) measures the timespan between two consecutive MOUSE_UP and MOUSE_DOWN events.

DISTANCE BETWEEN CLICKS
UNITS - pixels
Distance between clicks (DBC) 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
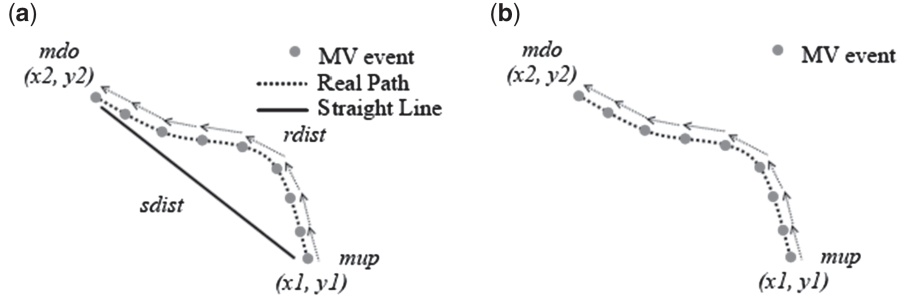
F$_{IG}$. 2. (a) A series of MOV events, between two consecutive clicks of the mouse. The difference between the shortest distance (sdist) and distance actually travelled by the mouse (rdist) is depicted. (b) The real distance travelled by the mouse between each two consecutive clicks is given by summing the distances between each two consecutive MOV events.

distance travelled by the mouse is given by (3).

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

DISTANCE OF THE MOUSE TO THE STRAIGHT LINE
U$_{NITS}$ - pixels
Distance of the mouse to the straight line (DMSL) 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 (2). However, it returns this sum rather than the average value during the path.

EXCESS OF DISTANCE
U$_{NITS}$ - pixels
Excess of distance (ED) 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
U$_{NITS}$ - pixels/milliseconds^2
Mouse acceleration (MA) is 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 MOUSE_UP and MOUSE_DOWN events, using the intervals and data computed for the Velocity.

MOUSE VELOCITY
U$_{NITS}$ - pixels/milliseconds

Mouse velocity (MV) is 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 *mdo*, respectively in the coordinates $(x1, y1)$ and $(x2, y2)$, that took place in the instants
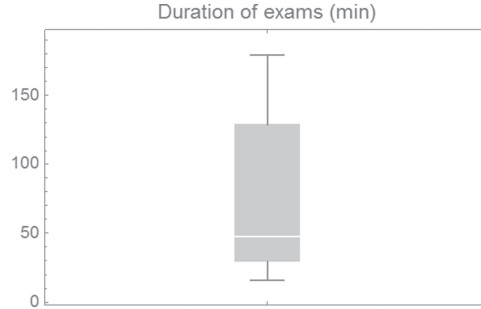
Duration of exams (min)

FIG. 3. Distribution of the duration (in minutes) of the 62 exams analysed.

$time_1$ and $time_2$, respectively. 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 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 (3).

TIME BETWEEN CLICKS (TBC)
UNITS - milliseconds
Time between clicks (TBC) is the timespan between two consecutive MOUSE_UP and MOUSE_DOWN events, i.e. how long did it take for the individual to perform another click.

4   Data analysis

As previously mentioned, data from 104 students was collected simultaneously. However, students were allowed to leave the exam at any moment. It is common for students to take a look at the exam for a few minutes before deciding to leave.

   To rule out these cases, we considered only data from students who had been, at least, 15 minutes at the computer. This resulted in a data set with data from 62 students. The duration of the exams is distributed as depicted in Figure 3 ($\bar{x} = 74.59$, $\tilde{x} = 47.54$, $Std = 53.28$).

   The features described in Subsection 3.2 were extracted for each student. In order to decrease the size of the data set, the values of the features were summarized at intervals of 2 minutes. The values of these features were plotted graphically, to visually assess their evolution for each student (Figure 4).

   A visual analysis of the data pointed out one first interesting aspect: in longer exams the curve that fitted the data seemed to be more flat. This is visible in the sample data in Figure 4, depicting three interaction features of student 39, who took 35 minutes to complete the exam, and student 42, who took 140 minutes.

   In order to determine if there was, indeed, a relationship between the shape of the curve and the duration of the exam, the following procedure was followed. For each student and each feature, a quadratic least-squares fit was found, which modeled the evolution of the value of the feature over time. Ten data sets were thus created. Taking into account that a quadratic function, in its standard form, is represented as $f(x) = ax^2 + bx + c$, each of these data sets contains the following attributes: studentID, c, b, a. Thus, each line of each data set (which details one feature) describes the curve
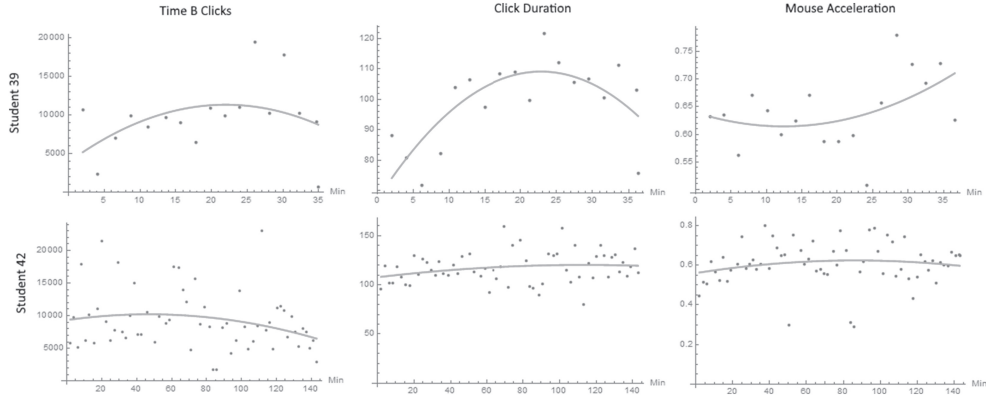
FIG. 4. Data of two students and three features, with curve fitting.

TABLE 1. Values of correlation between the coefficients *a* and *b* of the quadratic function and each of the features.

	ASA	ADMSL	AED	CD	DBC	DMSL	ED	MA	MV	TBC
a	−0.47	−0.42	−0.45	−0.64	−0.43	−0.45	−0.35	−0.64	−0.60	−0.52
b	−0.54	−0.42	−0.48	−0.66	−0.49	−0.44	−0.43	−0.6	−0.62	−0.51

that fits the values of a feature for a specific student. Since the duration of each exam is the same for each student regardless of the feature, a separate data set contains the studentID's and the duration of their exams.

We then calculated the correlation, for each feature, between the duration of the exam and the absolute values of *a* and *b*. The absolute value of the coefficients were used since we do not want to determine, for example, if the curve is convex or concave. We want to determine if it is more or less flat. The following values of correlation were obtained:

From the data depicted in Table 1, the results are clear that there is a generally significant negative correlation between the coefficients of the quadratic function and the duration of the exam. This correlation is stronger in features such as Click Duration, Mouse Acceleration and Mouse Velocity.

The negative values of correlation also support, in a generalized way, the initial claim that longer exams have a more flat curve. Indeed, in a quadratic function, the smaller the coefficient *a* the more flat the curve is. Figure 5 shows how this coefficient relates to the duration of the exam, for each feature.

5   Using Mouse Dynamics to predict exam duration

In this section we investigate the possibility of predicting the duration of the exam based on the way the student is interacting with the mouse. To this end, the data sets were randomly divided into 2 parts: one for training a model (containing 66% of the instances) and the other for validating it (containing the remaining 34% of the instances).

A Random Forest algorithm was used to train the predictor model. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [3].
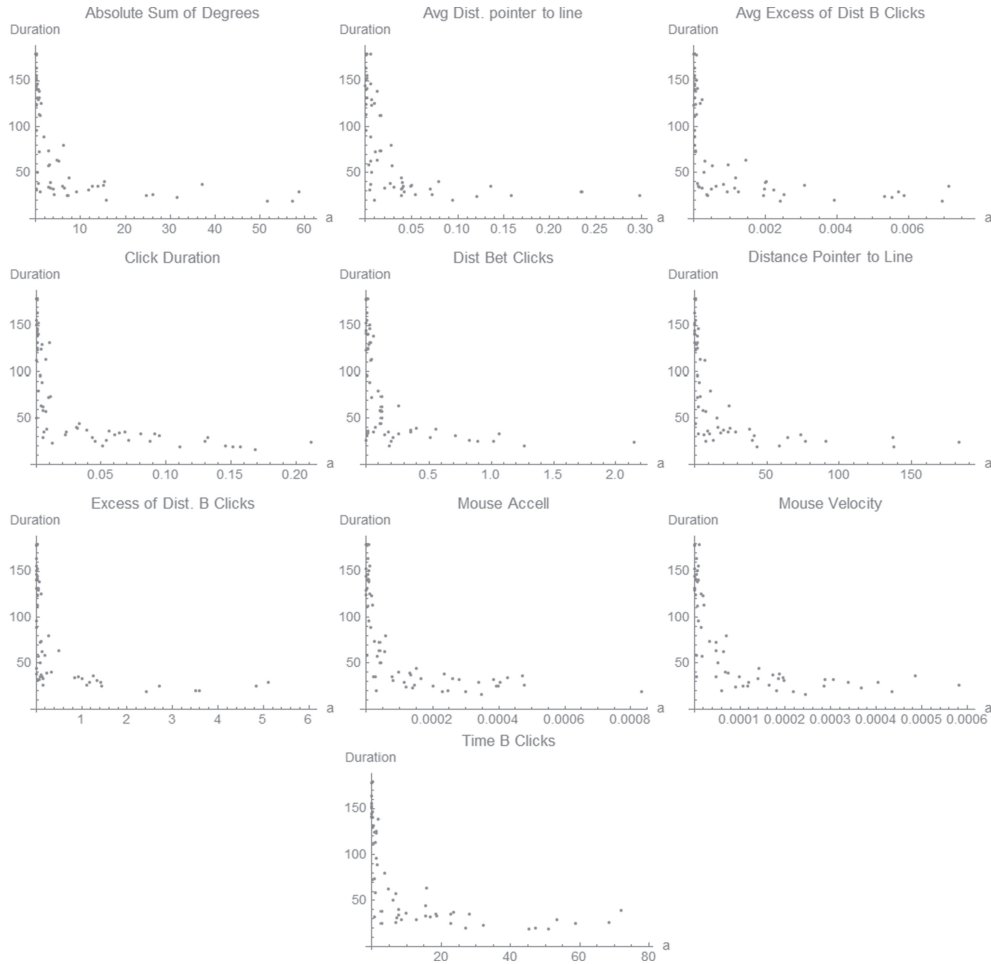
FIG. 5. Duration of each exam versus coefficient *a* of its quadratic fit, for each of the features studied.

The accuracy of the model was evaluated by predicting the duration of each of the instances in the validation set and computing the difference between the predicted duration and the actual duration. The accuracy is thus measured as the average error.

Since this accuracy depends, to some extent, on the randomized selection of the data, we repeated the process 20 times for each feature, with a random division of the data set (into training and validation) being performed in each iteration. Figure 6 depicts the distribution of the average errors, for each feature, during the 20 runs of this procedure. In some of the features, the minimum value of the error is nearly as low as 10 minutes. However, the best-average values of error are around 20 minutes, for the best-performing features.

It is also possible to visualize the relationship between the coefficients of the quadratic functions and the duration of the exam. Figure 7 depicts two surface plots. The one on the left depicts the
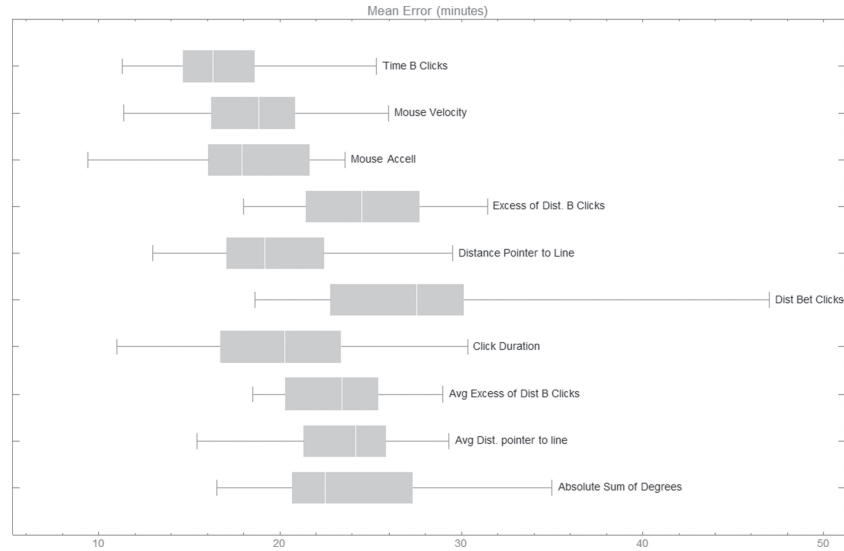
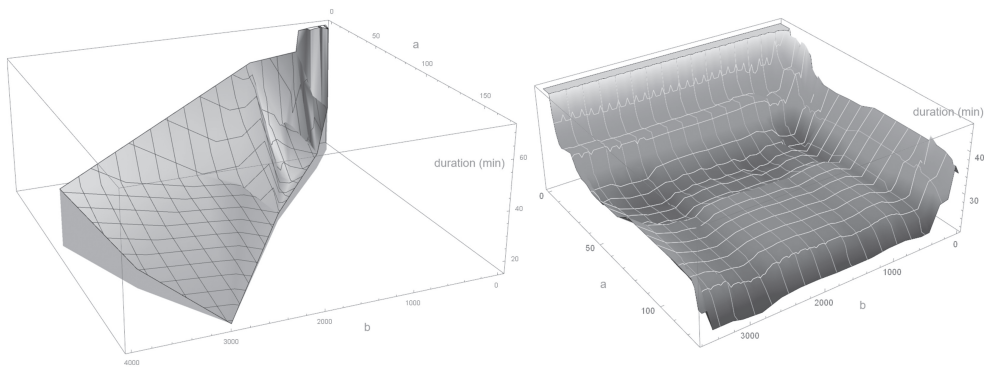FIG. 6. Distribution of the average error, for each feature, in 20 runs.

FIG. 7. Surface plots of the coefficients versus the duration of the exams (left) and of the coefficients versus the predicted duration of the exams (right).

relationship between these variables in the collected data. The one on the right uses the predictor model to predict the duration for the whole range of values of the coefficients *a* and *b*. Both point out the same inverse relationship between the coefficients and the duration of the exam.

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 different behavioural patterns can be observed during high-stakes computer-based exams. Namely, two specific behaviours can be clearly distinguished. There are students whose performance follows a steady trend (either increasing, decreasing or steady) and there are students whose performance during the exam follows a curve, either increasing and then decreasing or vice versa.

Secondly, this work also showed that this correlates strongly to the duration of the exam. That is, if we take the line that fits the temporal evolution of the student's performance, its shape (specifically its coefficients a and b) correlates to the duration of the exam.

Thirdly, we showed that this relationship is so strong that, depending on the features, we can predict the duration of the exam based on the interaction patterns with an average error as low as 15 minutes.

Our future work will, on the one hand, be guided by the aim of using this knowledge to useful ends. Although the prediction of the duration of the exam, in real-time, could be one of these aims, the utility in this may not be so significant. However, finding correlations of changing patterns of mouse dynamics with cognitive or work performance may be helpful in determining students under stress or in academic difficulty.

On the other hand, we will repeat this procedure to other cohorts of students in an attempt to replicate our findings and determine if they are similar for students with different ages and in different years of their course. We will also repeat this study using exams with different characteristics (e.g. relying more on the keyboard) as different and complementary conclusions will undoubtedly be reached in those cases.

The presented approach results are innovative in the sense that the data sources used have never been explored in the past with this purpose, i.e. to assess the effects of stress during high-stakes exams. For this reason, it may reveal previously unknown phenomena that may be used to more accurately characterize student's stress-coping mechanisms. Moreover, as opposed to traditional approaches, this system can be applied effortlessly, by installing a lightweight data-collecting software on the computer where student's are taking the exam. This allows for extensive and unobtrusive data collection procedures, that will allow for a better understanding of future healthcare professionals when under stress.

Acknowledgements

References

[1] A. A. E. Ahmedand and I. Traore. Anomaly intrusion detection based on biometrics. In *Proceedings of the IEEE Information Assurance Workshop*, pp. 452–453. West Point, NY, 2005.

[2] Y. Aksari and H. Artuner. Active authentication by mouse movements, In *Proceedings of the 24th International Symposium Computer and Information Science*, pp. 571–574. Guzelyurt, 2009.

[3] L. Breiman. Random forests. *Machine learning*, **45**, 5–2, 2001.

[4] S. M. Case and D. B. Swanson. Constructing written Test Questions for the Basic and Clinical Sciences. Philadelphia, PA: National Board of Medical Examiners, 1998.

[5] L. N. Dyrbye, M. R. Thomas and T. D. Shanafelt. Systematic review of depression, anxiety, and other indicators of psychological distress among US and Canadian medical students. *Academic Medicine*, **81**, 354–373, 2006.

[6] D. Goebert, D. Thompson, J. Takeshita, C. Beach, P. Bryson, K. Ephgrave, A. Kent, M. Kunkel, J. Schechter and J. Tate. Depressive symptoms in medical students and residents: a multischool study. *Academic Medicine*, **84**, 236–241, 2009.

[7] A. K. Jain, A. Ross and S. Prabhakar. An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, **14**, 4–20, 2004.

[8] R. Kohavi. Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, (KDD'96)*, Evangelos Simoudis, Jiawei Han and Usama Fayyad, eds, pp. 202–207. AAAI Press, 1996.

[9] B. S. McEwen. Brain on stress: how the social environment gets under the skin. *Proceedings of the National Academy of Sciences*, **109**, 17180–17185, 2012.

[10] A. Pimenta, D. Carneiro, P. Novais and J. Neves. Detection of distraction and fatigue in groups through the analysis of interaction patterns with computers. *Studies in Computational Intelligence*, vol. 570, pp. 29–39. Springer, 2015.

[11] C. Shen, Z. Cai, X. Guan, Y. Du and R. A. Maxion. User authentication through mouse dynamics. *IEEE - Information Forensics and Security*, **8**, 16–30, 2013.

[12] J. M. Soares, A. Sampaio, L. M. Ferreira, N. C. Santos, F. Marques, J. A. Palha, J. J. Cerqueira and N. Sousa. Stress-induced changes in human decision-making are reversible. *Translational psychiatry*, **2**, e131, 2012.

[13] R. D. Strous, N. Shoenfeld, A. Lehman, A. Wolf, L. Snyder and O. Barzilai. Medical students' self-report of mental health conditions. *International Journal of Medical Education*, **3**, 1, 2012.

[14] R. V. Yampolskiy and V. Govindaraju. Behavioural biometric: A survey and classification. *International Journal of Biometrics*, **1**, 81–113, 2008.