AgeGen Bio Track: Continuous Mouse Behavioral Biometrics-Based Age and Gender Profiling in Online Education Platforms

Aditya Subash[1][a], Insu Song[1][b], Ickjai Lee[2][c] and Kyungmi Lee[2][d]
[1]College of Science and Engineering, James Cook University, Singapore
[2]College of Science and Engineering, James Cook University, Cairns, Australia

Keywords: Mouse Behavioral Biometrics, Age Classification, Gender Classification, Online Education Platforms.

Abstract: Mouse behavioral biometric-based authentication systems have attracted significant attention as they are considered a more secure alternative to conventional online assessment fraud detection systems. This is attributed to their ability to continuously authenticate users non-intrusively by analyzing their distinctive mouse operating behavior. Most behavioral biometric-based research studies focus on predicting user identity as the primary objective for online assessment fraud detection. However, they do not consider predicting other user-centric parameters like age and gender. Furthermore, there is a need to identify the best segmentation approach and mouse behavior feature set for age and gender classification. We propose the AgeGen Bio track system, a continuous mouse behavioral biometric-based age and gender tracking system for online education platforms. To accomplish this, we first collect novel mouse behavior data with user demographic information. We then evaluate the efficacy of different segmentation approaches, feature sets, and machine learning models for age and gender classification. Experimental results show that the random forest algorithm paired with the three mouse-movement segmentation approach and user characteristic feature set are the best approaches that need to be incorporated into the system, as they achieved promising results.

1 INTRODUCTION

Recently, the education sector has advanced from offline to online settings over the past decade due to advancements in information technology (IT), resulting in widespread accessibility and proliferation of online education throughout the world (Garg and Goel, 2022). Several higher educational institutions (HEIs) now offer online courses as part of blended or fully online education (Garg and Goel, 2022; Wei et al., 2021). Like offline education, assessments are integral to any online educational curriculum and are often organized to evaluate learning outcomes, subject application, and knowledge retention (Garg and Goel, 2022). Despite the advantages of online assessments, a significant challenge is the prevalence of assessment fraud or academic dishonesty, often facilitated by the misuse of digital technologies (Blau and Eshet-Alkalai, 2017; Susnjak and McIntosh, 2024). These technologies make it straightforward to commit online assessment fraud by offering paid services and tools that assist students in credential sharing, fake identity matching, and plagiarism (Susnjak and McIntosh, 2024; Noorbehbahani et al., 2022). The current methods to prevent online assessment fraud, including conventional authentication approaches, are one-time, non-repudiable, intrusive, and expensive (Siddiqui et al., 2021; Subash et al., 2024).

Mouse behavioral biometric-based authentication systems have become a more popular and secure alternative to online assessment fraud detection. This is due to their cost effectiveness and ability to continuously verify user identity based on observable mouse movement behaviors (Zheng et al., 2011), which is inherently more challenging to spoof or replicate, making the approach more robust. Most behavioral biometric-based research studies focus on predicting user identity (ID) as the primary objective for security-related applications. However, they do not consider predicting other user-centric parameters, such as age and gender. Despite many studies in the field, there is no research and datasets for mouse behavioral biometric-based age and gender recognition

[a] https://orcid.org/0000-0003-0196-6256
[b] https://orcid.org/0000-0002-9835-0070
[c] https://orcid.org/0000-0002-6886-6201
[d] https://orcid.org/0000-0003-3304-4627

Subash, A., Song, I., Lee, I. and Lee, K.
AgeGen Bio Track: Continuous Mouse Behavioral Biometrics-Based Age and Gender Profiling in Online Education Platforms.
DOI: 10.5220/0013138500003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 3*, pages 383-393
ISBN: 978-989-758-737-5; ISSN: 2184-433X

in online education platforms. Furthermore, we must identify the best segmentation approach and mouse behavior feature set for our specific application.

We propose the AgeGen Bio track system, a continuous mouse behavioral biometric-based age and gender tracking system for online education platforms. To accomplish this, we first collect novel mouse behavior data with user demographic information. We then evaluate the efficacy of different segmentation approaches, feature sets, and machine learning models for age and gender classification. Specifically, we will evaluate several machine learning (ML) approaches, including Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), Support Vector Machine (SVM), Logistic regression (LR) and k-Nearest Neighbor (k-NN). The main contributions of the paper are as follows.

- An AgeGen Bio Track system capable of continuously monitoring user age and gender in online education platforms.
- Newly collected task-specific mouse behavior data with user-centric information from a case study for our experimentation.
- Comprehensive study comparing several segmentation approaches, feature sets, and machine learning models to identify the most effective approaches that can be integrated into the system.

2 BACKGROUND

This section will highlight the importance of including age and gender prediction in mouse behavioral biometrics-based online assessment fraud detection.

According to previous research (Van Balen et al., 2017), men and women differ in physical dimensions as described by anthropometric data recorded from several sources. These attributes can significantly impact the maneuverability of peripheral (mouse) devices across 2-dimensional surfaces, as it requires coordination between the arms, wrist, and fingers (Van Balen et al., 2017). Research conducted by (Fryar et al., 2012) indicates significant differences in arm lengths between males and females across several age groups. These variations could lead to differences in mouse movements (Van Balen et al., 2017).

Furthermore, several research studies regarding motor behavior indicate that men tend to move faster with less accuracy than women (Barral and Debû, 2004; Rohr, 2006). For example, (Rohr, 2006) conducted a study by requesting subjects to participate in a mouse-pointing task that required them to click targets of various sizes across the midline of a device. According to the study, women showed greater accuracy and slower deceleration time than men during the ballistic component of mouse movement (Van Balen et al., 2017).

Similarly, (Jiménez-Jiménez et al., 2011) also investigated the effect of age and gender on motor behavior. A total of 246 participants (123 males and 123 females) were recruited from seven age groups for the study. Participants were required to perform a set of physical and computer tasks. Parameters such as finger tapping frequency, movement time, walking time, and visual reaction time were measured for analysis. Results indicate that age and gender play a significant role in motor behavior. Furthermore, the speed of motor performance was found to be better in men (Jiménez-Jiménez et al., 2011).

Studies have also confirmed significant differences between typing speeds of different age groups due to generational differences. A study by (Pentel, 2017) confirmed that participants in the age group 16-19 were faster at typing than other age groups.

The aforementioned studies clearly indicate that age and gender characteristics are essential in understanding distinctive user behaviors, making them important parameters that must be included in current mouse behavioral biometric-based online assessment fraud detection systems.

3 RELATED WORK

This section will critically analyze, summarize, and present findings on several previous studies on behavioral biometric-based age and gender prediction. Specifically, we will summarize the several datasets, data collection strategies, features, and AI methodologies implemented in current research. For this purpose, we review several papers from well known publishers, including IEEE, Science Direct, and Springer.

Given the variety of behavioral biometrics studies, we will focus on reviewing articles related to age and gender classification using keystroke or mouse behavior biometrics.

3.1 Datasets and Data Collection Strategies

According to our review, datasets are classified into 1) Task-specific and 2) unconstrained datasets. Task-specific datasets collect user behavior based on predetermined mouse operation tasks (Zheng et al., 2011; Van Balen et al., 2017). Meanwhile, unconstrained datasets collect mouse behavior data by continuously monitoring users while they perform their daily activities without any constraints (Zheng et al., 2011; Van Balen et al., 2017). These types of data are collected to perform static or dynamic authentication.

We will now describe the various types of data used in research. This includes briefly describing public and novel datasets used in behavioral biometric-based age group and gender prediction.

3.1.1 Novel Datasets

1. Van Balen et al. (2017) collected mouse behavior data from 94 participants in a controlled environment for gender classification. Participants were required to perform a specific predetermined task involving identifying and clicking certain targets of different sizes located in 16 possible locations. Once the target is clicked, a new target will be displayed. Each participant performed several practice trials consisting of combinations of target size, target distance, and approach angle. After practice sessions, participants were then required to perform four blocks of 64 movement trials, with each block containing random sequences of two trials for each combination of target locations and sizes.

2. Tsimperidis et al. (2017, 2018, 2021) gathered unconstrained keystroke behavior data to predict age, gender, and educational qualifications (Tsimperidis et al., 2020). The data was captured using a key logger (IRecU) while participants engaged in their daily activities. Participants were also requested to provide demographic and educational details alongside the keystroke data. According to the authors, keystroke behavior data was collected from four age groups: 18-25, 26-35, 36-45, and 46+.

3. Tsimperidis et al. (2015) collected keystroke behavior data from 24 subjects using a key-logging application while participants were typing a fixed text of 850 characters, twice on a laptop and a desktop. In addition to keystroke behavior data, gender, and left/right-handedness information was also obtained (Tsimperidis et al., 2015).

4. Idrus et al. (2013, 2014) collected static keystroke behavior data from 110 participants for age, gender, and handedness classification. According to the study, participants belonged to two different countries, namely, France and Norway. Two keyboards with various keyboard layouts, AZERTY and QWERTY, are used during data collection. The data collection procedure involved participants writing five common phrases 20 times each. Furthermore, they were required to write the phrases ten times with one hand and ten times with two hands. According to the author, this data will be publicly available. Data acquisition was conducted using software developed using the publicly available GREYC software.

5. Pentel (2017, 2019) collected uncontrolled online key-stroke and mouse behavior data for age group and gender classification. User behavior data was collected from six sources, including the school's internal management system, feedback questionnaires, testing environments, and controlled experiments. Data was acquired using a JavaScript key-logging tool integrated into all six sources. According to the author, user behavior data was collected between 2011 and 2017 from several different age groups. Additional data such as screen resolution, device type (laptop, desktop, mobile devices), and operating system were also collected.

6. Kolakowska et al. (2016) collected keystroke and mouse behavior from 42 participants in a completely uncontrolled environment. Out of 42, 9 were females, and 33 were males. Data acquisition involved developing a browser plug-in that recorded relevant keystrokes and mouse behavior data while participants performed several activities in the browser. According to the study, different plug-in versions were made for Chrome and Opera browsers. Furthermore, user behavior data was collected from various age groups and peripheral devices (mouse, touchpad, trackpoint, touchpad).

3.1.2 Public Datasets

1. Buriro et al. (2016) used a publicly available keystroke dataset (TDAS) for age, gender, and operating hand estimation. The dataset collected keystroke data from 150 participants across six distinct age groups. Keystroke behavior data was collected while participants typed two PINs, one 4-digit (5560) and one 16-digit PIN (137966662480852), using a Samsung Galaxy Tablet. Furthermore, this dataset was collected in three different user-tagged location

2. Studies also implemented the publicly available GREYC keystroke behavior dataset for gender

Table 1: Datasets implemented by behavioral biometric-based age and gender research studies.

Author	Modality	Participants	Environment	Public	Type	Purpose
Van Balen et al., 2017	Mouse	94	Controlled	No	Task Specific	Gender Classification
Buriro et al., 2016	Keystroke	150	Semi-Controlled	Yes	Task Specific	Age/Gender/ Hand Classification
Tsimperidis et al., 2017	Keystroke	-	Uncontrolled	No	Unconstrained	Age Classification
Tsimperidis et al., 2018	Keystroke	75	-	No	Unconstrained	Gender Classification
Tsimperidis et al., 2015	Keystroke	24	Uncontrolled	No	Unconstrained	Gender Classification
Fairhurst and Da Costa-Abreu, 2011	Keystroke	133	Controlled	Yes	Task Specific	Gender Classification
Giot and Rosenberger, 2012	Keystroke	133	Controlled	Yes	Task Specific	Gender Classification
Idrus et al., 2014	Keystroke	110	Controlled	Yes	Task Specific	Age/Gender/ Hand Classification
Idrus et al., 2013	Keystroke	110	Controlled	Yes	Task Specific	Gender/Hand Classification
Pentel, 2017	Keystroke /Mouse	1519	Uncontrolled	Yes	Unconstrained	Age/Gender Classification
Tsimperidis et al., 2021	Keystroke	118	Uncontrolled	No	Unconstrained	Age/Gender Classification
Pentel, 2019	Keystroke	7119	Uncontrolled	Yes	Unconstrained	Age Classification
Kolakowska et al., 2016	Keystroke /Mouse	42	Uncontrolled	No	Unconstrained	Gender Classification

prediction. This dataset contains data from 133 participants who were required to type a predetermined password several times. In addition to keystroke data, gender data was also acquired. Among 133 participants, 98 are male and 35 are female (Fairhurst and Da Costa-Abreu, 2011; Giot and Rosenberger, 2012).

From our review (Table 1), it is evident that research on keystroke behavior biometrics has seen a marked increase, particularly in the context of predicting age and gender, in comparison to mouse behavioral biometrics. Furthermore, there are very few publicly available datasets for mouse behavioral biometric-based age and gender prediction.

3.2 Behavior Biometric Features

We will discuss the various features employed in behavioral biometric-based age and gender classification. Given that our background review encompasses multiple biometric modalities, we will categorize and emphasize the features according to each modality. Specifically, we will focus on keystroke and mouse behavior-based features for age and gender prediction.

3.2.1 Keystroke Behavior Features

After data collection, raw data such as key code (Tsimperidis et al., 2017; Tsimperidis et al., 2018; Syed Idrus et al., 2014; Idrus et al., 2013; Pentel, 2017; Pentel, 2019), action type (key press or release) (Tsimperidis et al., 2017; Tsimperidis et al., 2018; Syed Idrus et al., 2014; Idrus et al., 2013; Pentel, 2017; Pentel, 2019; Kolakowska et al., 2016), the date the action took place, and timestamp are collected and (Tsimperidis et al., 2017; Tsimperidis et al., 2018; Syed Idrus et al., 2014; Idrus et al., 2013; Pentel, 2017; Pentel, 2019; Kolakowska et al., 2016) are used for feature extraction.

Using this data, dwell (hold) time (Buriro et al.,

Table 2: AI-based approaches used in behavioral biometric-based age and gender classification.

Author	DL	ML	Fusion	DBC
Van Balen et al., 2017	No	LS-MR + LR	No	No
Buriro et al., 2016	NN, DNN	NB, SVM, RF	No	No
Tsimperidis et al., 2017	ANN (MLP)	OneR, SVM, BFDT, DT, NB, SL	No	No
Tsimperidis et al., 2018	MLP, RBFN	SVM, RF, NB	No	No
Tsimperidis et al., 2015	No	NB, SVM, RF	No	Manhattan, Euclidean
Fairhurst and Da Costa-Abreu, 2011	No	KNN, NB, DT	DCS-LA, Majority Voting, Sum	No
Giot and Rosenberger, 2012	No	SVM	Score, Template	No
Idrus et al., 2014	No	SVM	Majority Voting, Score	No
Idrus et al., 2013	No	SVM	No	No
Pentel, 2017	No	LR, SVM, KNN, DT, RF	No	No
Tsimperidis et al., 2021	RBFN	SVM, SL, NB, BNC	No	No
Pentel, 2019	No	SVM, RF, FT, LR	No	No
Kolakowska et al., 2016	NN	BNC, RT, DT, AdaBoost	No	No

2016; Fairhurst and Da Costa-Abreu, 2011; Pentel, 2017; Tsimperidis et al., 2021), release-press latency (Buriro et al., 2016; Fairhurst and Da Costa-Abreu, 2011; Giot and Rosenberger, 2012; Syed Idrus et al., 2014; Pentel, 2017), release-release latency (Buriro et al., 2016; Fairhurst and Da Costa-Abreu, 2011; Giot and Rosenberger, 2012; Syed Idrus et al., 2014), press-release latency (Buriro et al., 2016; Fairhurst and Da Costa-Abreu, 2011; Giot and Rosenberger, 2012; Syed Idrus et al., 2014), press-press latency (Buriro et al., 2016; Fairhurst and Da Costa-Abreu, 2011; Giot and Rosenberger, 2012; Syed Idrus et al., 2014; Tsimperidis et al., 2021), and a vector combination of all the previous latencies (Giot and Rosenberger, 2012; Syed Idrus et al., 2014; Idrus et al., 2013), are extracted and used for age and gender classification.

Detailed analysis also reveals that studies frequently extract the most common diagram patterns and compute the average and standard deviation values of n-graph-latencies and hold times for analysis (Tsimperidis et al., 2017; Tsimperidis et al., 2018; Pentel, 2017; Pentel, 2019). Additional features extracted include the relative frequency of corrective keys (DEL) (Pentel, 2017; Kolakowska et al., 2016), average time (Pentel, 2017), correctness (ratio between the number of keystrokes and the number of characters in the final text) (Pentel, 2017), percentage use of special character keys (Kolakowska et al., 2016), pauses between words (Pentel, 2019), and statistical measurements of (average, standard deviation, maximum, minimum, variance, mode, and range) of typing speed, hold time, and press-press latency were also computed (Kolakowska et al., 2016).

Studies were also found to integrate gender classification into keystroke authentication systems for enhanced performance (Giot and Rosenberger, 2012). Feature selection algorithms, such as Information Gain (IG), and oversampling techniques, including SMOTE, have also been implemented for analysis (Buriro et al., 2016; Tsimperidis et al., 2018; Tsimperidis et al., 2021).

3.2.2 Mouse Behavior Features

Raw mouse behavior data, such as timestamp (Van Balen et al., 2017; Pentel, 2017), coordinates space (x, y) (Van Balen et al., 2017; Pentel, 2017), action type (Van Balen et al., 2017; Kolakowska et al., 2016), and location/size of the target (Van Balen et al., 2017), are recorded and used for feature extraction.

Using this data, several different types of mouse behavior features, such as temporal, spatial, and accuracy metrics, were calculated (Van Balen et al., 2017).

These metrics can again be subdivided into several types of features, including reaction time (RT), peak velocity (PK), time to peak velocity (TPV), duration of ballistic movement (DB), shape of velocity profile (SV), proportion of ballistic movement (PB), number of movement corrections (NC), time to click (TC), hold time (HT), movement time (MT), path length (PL), path length to best path ratio (PLR), task axis crossings (TXC), movement direction changes (MDC), orthogonal movement changes (MDC), movement variability (MV), absolute error (AE), horizontal error (HE), vertical error (VE), absolute horizontal error (AHE), and absolute vertical error (AVE) (Van Balen et al., 2017). Additional attributes such as distance, angle, velocity, movement, acceleration, action, and direction-based features were also implemented by current research studies (Pentel, 2017; Kolakowska et al., 2016).

3.3 Machine Learning Approaches

This section highlights the several AI methods used in mouse behavioral biometric-based age and gender analysis. On analysis, we find that approaches, including logistic regression (LR) (Pentel, 2017; Pentel, 2019), support vector machine (SVM) (Buriro et al., 2016; Tsimperidis et al., 2017; Tsimperidis et al., 2018; Tsimperidis et al., 2015), random forest (RF) (Buriro et al., 2016; Tsimperidis et al., 2018; Pentel, 2017), k-nearest neighbors (KNN) (Fairhurst and Da Costa-Abreu, 2011; Pentel, 2017), OneR (Tsimperidis et al., 2017), best first decision tree (BFDT) (Tsimperidis et al., 2017), rotation forest (RT) (Kolakowska et al., 2016), AdaBoost (Kolakowska et al., 2016), simple logistics (SL) (Tsimperidis et al., 2017; Tsimperidis et al., 2021), decision tree (DT) (Tsimperidis et al., 2017; Tsimperidis et al., 2015; Fairhurst and Da Costa-Abreu, 2011; Pentel, 2017), Bayesian network classifier (BNC) (Tsimperidis et al., 2021; Kolakowska et al., 2016), and naïve Bayes (NB) (Buriro et al., 2016; Tsimperidis et al., 2017; Tsimperidis et al., 2018; Tsimperidis et al., 2015) are some of the popular ML approaches implemented in the field. A few studies also use a combination of ML approaches for classification. For example, Van Balen et al. (2017) implements a combination of least-squares multiple regression (LS-MR) and LR for gender classification.

Upon further analysis, several studies also utilize distance-based classifiers (DBC) and various fusion techniques (Tsimperidis et al., 2015; Fairhurst and Da Costa-Abreu, 2011), which include Manhattan distance, Euclidean distance, dynamic classifier Selection based on local accuracy (DCS-LA), majority voting, sum-based methods, template information, and score-based fusion approaches (Tsimperidis et al., 2015; Fairhurst and Da Costa-Abreu, 2011; Giot and Rosenberger, 2012; Syed Idrus et al., 2014).

Despite the popularity of ML approaches, deep learning (DL) approaches were also found to be implemented in the field. For example, (Buriro et al., 2016; Kolakowska et al., 2016) both implement a novel neural net (NN) and deep neural network architecture (DNN) for behavioral biometric-based user characteristics (age, gender, or operating handedness) classification. Similarly, (Tsimperidis et al., 2017; Tsimperidis et al., 2018) implements MLP for keystroke behavioral biometric-based age and gender prediction. In addition to the conventional DL approaches applied previously, radial basis function networks (RBFN) (Tsimperidis et al., 2018; Tsimperidis et al., 2021) have also been implemented. Further investigation shows that the research studies also rely on meta-algorithms, such as AdaBoost, multiboot, random-correction-code, exhaustive-correction-code, and rotation forest, to boost classifier performance (Tsimperidis et al., 2018; Tsimperidis et al., 2021).

Table 2 summarizes the AI approaches and evaluation criteria (EC) implemented for analysis. Based on the information presented in Table 2, we find that the ML approaches are more popular for behavioral biometric-based age and gender classification despite the availability of advanced DL approaches (Table 2). Further analysis shows that SVM is the most popular ML approach for behavioral biometric-based age and gender prediction.

4 DATA COLLECTION PROCEDURE

Our background analysis confirms that there are no publicly available mouse behavior datasets for continuous age and gender classification in online education platforms (Table 1). Hence, we collect novel mouse behavior data for our specific application.

Before we pursue data collection, we need to understand what type of mouse behavior data needs to be collected for accurate age and gender classification. This is realized through our comprehensive

background analysis, which reveals that prior re-

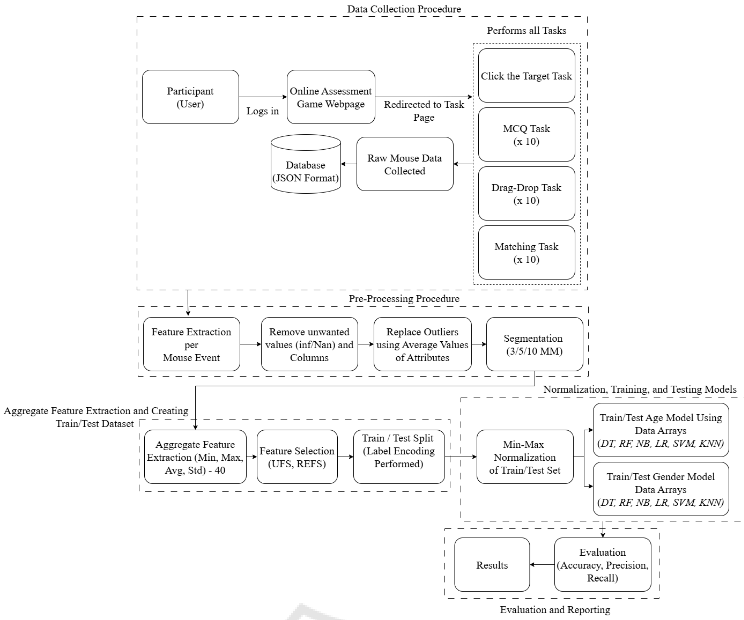

Figure 1: Methodology used in our study for mouse behavioral biometric-based age and gender classification.

search gathers mouse behavior data by instructing participants to complete predefined tasks in a controlled environment or by observing their daily activities in an uncontrolled setting (Van Balen et al., 2017; Pentel, 2017; Pentel, 2019).

Since our primary objective is to perform age and gender classification in online education platforms, we find that the former data collection method is more suitable for our experimentation. Hence, we collect novel mouse behavior data while participants engage in an online assessment game, which consists of several assessment-like tasks, including clicking the target, MCQ, drag-drop, and matching tasks.

The rationale for incorporating various tasks in the data collection procedure is derived from our background analysis. Our investigation shows that several mouse behavioral biometric-based authentication studies collect several varieties of mouse event data (Siddiqui et al., 2021; Subash et al., 2024; Zheng et al., 2011), including mouse movements, clicks, drag, and scroll events for analysis. Therefore, we included four simple but different tasks in our data collection procedure to collect the same variety of mouse behavior data.

Data acquisition was done with the help of a web application developed using HTML, CSS, and JavaScript. Necessary data is recorded using the mouse event listener method. In addition to this, we also collect screen dimension information using existing JavaScript methods. The description of the tasks is as follows:

1. MCQ Task: Contains four simple general knowledge questions that participants answer by selecting the correct choice among four choices. The subsequent question is displayed only when the current question is answered correctly. Once a choice is selected, the participant must click the submit button. If the choice is incorrect, the participants are shown a prompt indicating they must answer the question again. In other words, the user can rectify their answer until the correct choice is selected.

2. Click the Target Task: This task involves clicking a target (button element) that alters its position each time it is clicked. In total, the target changes position nine times.

3. Drag-Drop Task: Requires participants to drag an image of an animal into the correct drop-box containing the label of the animal's category (mammal, amphibian, reptile, fish, or bird). There are five images and five drop-boxes. If the participant drags the image in-to the correct dropbox, the background is changed to green, indicating a proper response. Furthermore, the scaled version of the image will be displayed within the box. If the participant drags an image into the in-correct dropbox, the images are returned to their original positions, and the background is changed to red for a short time frame. After which, the background color returns to its original state. Similarly

Table 3: Comparison of 3-, 5-, and 10-MM segments for age classification.

Segmentation Method	Approaches	Acc (%)	Pre (%)	Rec(%)
3-MM	SVM	56.41	65.61	50.61
	LR	57.57	56.99	53.26
	NB	56.95	55.96	52.37
	KNN (k= 6)	63.06	62.77	60.80
	RF*	76.61	76.46	75.83
	DT	67.76	67.41	66.44
5-MM	SVM	56.24	61.96	51.19
	LR	53.40	52.25	52.15
	NB	49.03	51.42	51.20
	KNN (k= 6)	59.28	58.52	57.23
	RF	59.89	61.63	61.19
	DT	55.43	57.28	56.87
10-MM	SVM	56.96	61.94	54.12
	LR	54.94	54.21	53.61
	NB	54.94	54.31	53.01
	KNN (k= 6)	59.19	59.89	57.46
	RF	54.34	56.92	55.80
	DT	55.95	56.29	56.28

to MCQ tasks, the participants are allowed to correct their mistakes.

4. Matching Task: This is the final task that the participants perform. They must identify four pairs of matching images (country flags) among eight images displayed on the screen. If the selected images do not match, they are shown briefly and restored to their original state.

Data was collected from 20 participants recruited from Sanjay Gandhi College of Education, Bengaluru, India. As part of the data collection process for this study, all participants were required to provide informed consent prior to their involvement. The process was carefully designed to ensure compliance with ethical guidelines and maintain transparency regarding the nature of the research and the use of collected data.

All participants were required to perform MCQ, drag-drop, and matching tasks ten times each. In addition to mouse behavior data, we collected user demographic (age group and gender) information via a pre-participation questionnaire. Among the 20 participants, ten are female and ten are male, distributed across two age groups: 18-22 and 23-27 years of age. During task engagement, raw data, such as timestamp, screen height, screen width of the content area, screen coordinates (X, Y), event action types, element on which the event was performed, offset X, and Y are collected for further feature extraction. Mouse behavior data is received individually for each user and task type in JSON format.

5 EXPERIMENTAL RESULTS

This section will present the experimental results, using which we identify the best segmentation method, ML model, and features set to be integrated into the AgeGen Bio Track system for continuous age and gender classification.

As mentioned before, we will compare several different ML models, including DT, RF, NB, LR, SVM, and kNN. Training and testing will be performed using the hold-out approach, where 90% will be used for training, and 10% will be used for testing. Before performing segmentation, model training, and evaluation, we will extract features per mouse event and remove any unwanted values and columns. We then pre-process the data by replacing outliers using the average values of the attributes and performing necessary segmentation methods. The overall working methodology is illustrated in Figure 1.

5.1 Best Segmentation Method and ML Model for Age and Gender Classification

Before training the ML models, we apply segmentation as a preprocessing technique to logically group mouse behavior data into meaningful blocks of information. These segments are then used to compute aggregate values of the attributes, such as minimum (min), maximum (max), average (Avg), and standard deviation (std) (Subash et al., 2024). For our experimentation, we will perform segmentation according

Table 4: Comparison of 3-, 5-, and 10-MM segments for gender classification.

Segmentation Method	Approaches	Acc (%)	Pre (%)	Rec(%)
3-MM	SVM	58.79	59.45	58.45
	LR	58.54	58.53	58.54
	NB	55.12	58.49	54.41
	KNN (k= 6)	62.14	62.57	61.91
	RF*	75.76	75.75	75.76
	DT	69.41	69.51	69.47
5-MM	SVM	52.58	54.05	53.14
	LR	53.60	56.66	54.30
	NB	55.12	59.15	54.27
	KNN (k= 6)	57.36	57.42	57.13
	RF	50.55	53.72	51.46
	DT	54.82	54.78	54.78
10-MM	SVM	59.79	60.36	60.27
	LR	58.78	58.53	57.69
	NB	52.72	55.21	54.26
	KNN (k= 6)	59.39	61.10	60.38
	RF	52.52	49.82	49.91
	DT	55.15	54.85	54.84

to the procedure mentioned in (Subash et al., 2024), which uses the n-mouse movement (MM) segmentation approach. This approach groups a predefined number (n) of mouse movement events into a single segment for analysis (Siddiqui et al., 2021; Subash et al., 2024).

However, based on our background analysis, the n-value varies according to the study performed (Siddiqui et al., 2021; Subash et al., 2024). In other words, there is no standard n value for analysis. Therefore, to understand the effect and identify the best MM segmentation approach, we test n in three scenarios: 3, 5, and 10. For evaluation, we extracted 40 features identified in our previous work (Subash et al., 2024) using data only from the first attempt. After feature extraction, we perform a train-test split (90/10), min-max normalization, and label encoding.

Based on the results (Tables 3 and 4), we conclude that 3-MM segmentation is the best-performing segmentation approach for our specific application. Furthermore, RF is the best performing ML algorithm evaluated under the 3-MM segmentation approach, achieving more than 75% accuracy, precision, and recall. It is also observed that SVM, DT, RF, LR, NB, and KNN achieve similar performance when implementing the 5- and 10-MM segmentation approach. A significant increase in performance is detected in RF and DT when we compare 3-, 5- and 10-MM segmentation approaches.

5.2 Identifying Best Feature Set for Age and Gender Classification

In the previous section, we identified the best-performing segmentation approach and ML model. Our experimentation confirms that the 3-MM segmentation approach evaluated on the RF model outperforms other approaches considered in this study.

In this section, we will improve the performance of the RF model by implementing feature selection approaches. Specifically, we will compare two approaches, univariate and recursive elimination feature selection (FS) algorithms, to determine the best feature set for age and gender classification. For our study, we perform FS after Min-Max normalization as the attributes have different SI units depending on the feature extracted. Furthermore, we selected the top 20 features for our analysis.

Table 5: Comparison of different FS approaches for age classification.

ML	FS Approach	Acc (%)	Pre (%)	Rec (%)
RF	Original Feature Set	76.61	76.46	75.83
RF	Univariate FS	72.71	72.62	71.55
RF	Recursive Elimination FS	80.76	80.83	79.98

From this experiment, it can be confirmed that the RFE-selected feature set (Table 7) is suitable for mouse behavioral biometric-based age and gender classification in online education platforms. Based on the performance achieved (Tables 5 and 6), it is evident that the recursive feature elimination approach gives the best feature set for age and gender classification. It is noticed that there is a 4-5% increase in performance in all evaluation criteria when we compare the performance between the RFE-selected features and the original feature set.

Table 6: Comparison of different FS approaches for gender classification.

ML	FS Approach	Acc (%)	Pre (%)	Rec (%)
RF	Original Feature Set	75.76	75.75	75.76
RF	Univariate FS	72.95	72.94	72.94
RF	Recursive Elimination FS	79.91	79.91	79.92

Table 7: User characteristic feature set selected for age and gender classification.

Aggregate Values	Attributes
Average (Avg)	Co-ordinates (x, y), timestamp, angle to path tangent, jerk
Minimum (Min)	Co-ordinates (x, y), vertical velocity, horizontal velocity, acceleration, timestamp
Maximum (Max)	Co-ordinates (x, y), vertical velocity, acceleration, jerk, distance, angular velocity, timestamp
Stand-alone value	Elapsed Time

6 CONCLUSION

It is essential to identify user characteristics, such as age and gender, to improve online assessment fraud detection systems that safeguard online education platforms. This paper proposes AgeGen Bio Track: a continuous mouse behavioral biometric-based age and gender classification model for this purpose. To accomplish this, we collected novel task-specific mouse behavior data while participants engaged in an online assessment game, using which we identified the best segmentation approach, machine learning model, and feature set for continuous age and gender classification. Our investigation indicates that our segmentation approach with the RF algorithm and user characteristic feature set attains satisfactory performance of 80% in all evaluation criteria. The overall performance achieved by our proposed approach indicates positive results in mouse behavioral biometric-based age and gender classification for our specific application. Our findings and comprehensive background analysis also support further research in the field and suggest that user age and gender parameters can be fused for behavioral biometric-based authentication to enhance performance.

REFERENCES

Barral, J. and Debû, B. (2004). Aiming in adults: Sex and laterality effects. *Laterality (Hove)*, 9(3):299–312.

Blau, I. and Eshet-Alkalai, Y. (2017). The ethical dissonance in digital and non-digital learning environments: Does technology promotes cheating among middle school students? *Computers in human behavior*, 73:629–637.

Buriro, A., Akhtar, Z., Crispo, B., and Del Frari, F. (2016). Age, gender and operating-hand estimation on smart mobile devices. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. Gesellschaft fuer Informatik.

Fairhurst, M. and Da Costa-Abreu, M. (2011). Using keystroke dynamics for gender identification in social network environment. In *4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011)*, pages P27–, Stevenage. IET.

Fryar, C. D., Gu, Q., and Ogden, C. L. (2012). Anthropometric reference data for children and adults: United states, 2007-2010. *Vital and health statistics. Series 11. Data from the National Health Survey*, (252):1–48.

Garg, M. and Goel, A. (2022). A systematic literature review on online assessment security: Current challenges and integrity strategies. *Computers & Security*, 113:102544.

Giot, R. and Rosenberger, C. (2012). A new soft biometric approach for keystroke dynamics based on gender recognition. *International Journal of Information Technology and Management*, 11(1-2):35–49. PMID: 44062.

Idrus, S. Z. S., Cherrier, E., Rosenberger, C., and Bours, P. (2013). Soft biometrics database: A benchmark for keystroke dynamics biometric systems. In *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*, pages 1–8. Gesellschaft für Informatik e.V. (GI).

Jiménez-Jiménez, F. J., Calleja, M., Alonso-Navarro, H., Rubio, L., Navacerrada, F., Pilo-de-la Fuente, B., Plaza-Nieto, J. F., Arroyo-Solera, M., García-Ruiz, P. J., García-Martín, E., and Agúndez, J. A. (2011). Influence of age and gender in motor performance in healthy subjects. *Journal of the neurological sciences*, 302(1):72–80.

Kolakowska, A., Landowska, A., Jarmolkowicz, P., Jarmolkowicz, M., and Sobota, K. (2016). Automatic recognition of males and females among web browser users based on behavioural patterns of peripherals usage. *Internet research*, 26(5):1093–1111.

Noorbehbahani, F., Mohammadi, A., and Aminazadeh, M. (2022). A systematic review of research on cheating in online exams from 2010 to 2021. *Education and information technologies*, 27(6):8413–8460.

Pentel, A. (2017). Predicting age and gender by keystroke dynamics and mouse patterns. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, pages 381–385, New York, NY, USA. ACM.

Pentel, A. (2019). Predicting user age by keystroke dynamics. In Silhavy, R., editor, *Artificial Intelligence and Algorithms in Intelligent Systems*, pages 336–343, Cham. Springer International Publishing.

Rohr, L. E. (2006). Gender-specific movement strategies using a computer-pointing task. *Journal of motor behavior*, 38(6):431–137.

Siddiqui, N., Dave, R., and Seliya, N. (2021). Continuous authentication using mouse movements, machine learning, and minecraft. *arXiv.org*.

Subash, A., Song, I., Lee, I., and Lee, K. (2024). Mouse dynamics-based online fraud detection system for online education platforms. In Yang, X.-S., Sherratt, S., Dey, N., and Joshi, A., editors, *Proceedings of Ninth International Congress on Information and Communication Technology*, pages 257–269, Singapore. Springer Nature Singapore.

Susnjak, T. and McIntosh, T. (2024). Chatgpt: The end of online exam integrity? *Education sciences*, 14(6):656–.

Syed Idrus, S. Z., Cherrier, E., Rosenberger, C., and Bours, P. (2014). Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords. *Computers & security*, 45:147–155.

Tsimperidis, I., Arampatzis, A., and Karakos, A. (2018). Keystroke dynamics features for gender recognition. *Digital investigation*, 24:4–10.

Tsimperidis, I., Katos, V., and Clarke, N. (2015). Language-independent gender identification through keystroke analysis. *Information and computer security*, 23(3):286–301.

Tsimperidis, I., Rostami, S., and Katos, V. (2017). Age detection through keystroke dynamics from user authentication failures. *International journal of digital crime and forensics*, 9(1):1–16.

Tsimperidis, I., Yoo, P. D., Taha, K., Mylonas, A., and Katos, V. (2020). R2bn: An adaptive model for keystroke-dynamics-based educational level classification. *IEEE transactions on cybernetics*, 50(2):525–535.

Tsimperidis, I., Yucel, C., and Katos, V. (2021). Age and gender as cyber attribution features in keystroke dynamic-based user classification processes. *Electronics (Basel)*, 10(7):835–.

Van Balen, N., Ball, C., and Wang, H. (2017). Analysis of targeted mouse movements for gender classification. *EAI endorsed transactions on security and safety*, 4(11):153395–15.

Wei, X., Saab, N., and Admiraal, W. (2021). Assessment of cognitive, behavioral, and affective learning outcomes in massive open online courses: A systematic literature review. *Computers and education*, 163:104097–.

Zheng, N., Paloski, A., and Wang, H. (2011). An efficient user verification system via mouse movements. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 139–150, New York, NY, USA. ACM.