Subash *et al.*
EURASIP Journal on Information Security (2025) 2025:21
https://doi.org/10.1186/s13635-025-00207-5

EURASIP Journal on Information Security

RESEARCH

Open Access

Integrating user demographic parameters for mouse behavioral biometric-based assessment fraud detection in online education platforms

Aditya Subash[1*], Insu Song[1], Ickjai Lee[2] and Kyungmi Lee[2]

Abstract

Online education systems have gained immense popularity due to their ubiquity, flexibility, openness, and accessibility. This has led many higher education institutions to incorporate online courses as part of blended or fully online learning. However, online assessment fraud remains a critical challenge. Conventional assessment fraud detection methods are often one-time, non-repudiable, invasive, expensive, and susceptible to spoofing. Even some advanced systems based on behavioral biometrics report comparatively lower accuracy, underscoring the ongoing challenge of achieving reliable user authentication. Furthermore, few research studies focus on behavioral biometric-based assessment fraud detection in online education platforms. To address these gaps, we introduce the UserID.AGE.GEN framework, which implements a cross-referencing fusion algorithm that integrates user demographic parameters, including age and gender, with mouse behavioral biometrics for user identity verification for online assessment fraud. Additionally, we collect novel task-specific data for our evaluation. Experimental results demonstrate that our method achieves promising results compared to some existing models, highlighting its strong performance and promising potential for broader application and future enhancement. A notable limitation of the proposed model is that it has not yet been evaluated using significantly larger external datasets, which may affect the generalizability of the results. Our evaluation was conducted using internally collected datasets. Additionally, the model has not been tested in real-world settings such as online education platforms, which may limit insights into its practical deployment.

Keywords Online assessment fraud detection, High false positive rates, UserID.AGE.GEN framework, Cross-referencing fusion algorithm, Online education platforms

1 Introduction

Online education refers to teaching and facilitating learning through digital technologies and platforms, such as Zoom, Microsoft Teams, and Moodle [1–3]. Recently, online education has gained immense popularity due to its ubiquity, flexibility, openness, and reach, which is usually attributed to the advancements made in information and communication technologies (ICT) [2]. Due to these advancements, several higher educational institutions (HEIs) now offer online courses as part of blended learning or fully online education [2, 4]. According to recent

*Correspondence:
Aditya Subash
aditya.subash@my.jcu.edu.au
[1] College of Science and Engineering, James Cook University, Singapore, Singapore
[2] College of Science and Engineering, James Cook University, Cairns, Australia

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

research, Moodle is the most commonly implemented online education platform, with estimates showing that more than 30,000 educational institutions utilize the tool for communication, learning, and consultation [2].

Like offline education, assessments are integral to any online education platform as they ensure an accurate evaluation of learning objectives, knowledge retention, and subject application [4]. Despite the advantages of online education, there are significant vulnerabilities, such as online assessment fraud or academic dishonesty [4]. Online assessment fraud is unethical behavior that violates fairness and integrity [5], often facilitated by the misuse of digital technologies that assist students in credential sharing, fake identity matching, and plagiarism [5, 6]. Typical online assessment fraud detection methods like authentication systems and online proctoring services are one-time, non-repudiable, intrusive, expensive, and vulnerable to spoofing [3, 6, 7]. Therefore, it becomes imperative for the research community to devise inexpensive, non-intrusive, and robust online assessment fraud detection systems capable of continuously ascertaining user identity in online education platforms.

Recently, mouse behavioral biometrics-based authentication systems have been considered a more secure alternative to conventional online assessment fraud detection systems, as they are more inexpensive, robust, and forgery-resistant [8]. They identify users by analyzing their unique behaviors while interacting with standard hardware devices such as mouse, screens, and trackpads [7]. Despite the volume of studies in the field, most research studies focus on predicting user identity (user ID) as a stand-alone process and suffer from high false positive rates. Furthermore, few research studies focus on mouse behavioral biometric-based authentication in online education platforms.

To address the identified problems, we propose an innovative UserID.AGE.GEN framework, which implements a novel cross-referencing fusion algorithm that integrates user demographic parameters, such as age and gender, to enhance user identity verification for online assessment fraud detection. Specifically, the proposed method will combine the precited decisions from the user ID, age group, and gender models by cross-referencing the predicted users with their age group and gender information to consolidate the decisions made by the user ID prediction model. We also acquire novel task-specific mouse behavior data for our evaluation. We will evaluate the combination of several machine learning (ML) approaches to identify the highest-performing model, which will be implemented for the cross-referencing fusion model. Specifically, we will be comparing random forest (RF), naïve bayes (NB), *K*-nearest neighbors (*K*-NN), light gradient boosting machine (LGBM), and multi-layered perceptron (MLP) for user ID, age, and gender prediction. This evaluation also helps determine the best models to combine in the cross-referencing algorithm.

The main contributions of the paper are as follows:

- Proposing an innovative UserID.AGE.GEN framework implementing a novel cross-referencing fusion algorithm that integrates user demographic parameters to enhance online assessment fraud detection.
- Collecting task-specific mouse behavior data with user demographic information from a case study for our experimentation.
- Undertaking comprehensive experimentation evaluating several segmentation methods, feature sets, and ML algorithms, such as RF, *K*-NN, MLP, NB, and LGBM for stand-alone age, gender, and user-id prediction models.
- Comparing our proposed novel cross-referencing fusion algorithm with models previously published.

2 Age and gender prediction using behavioral biometrics

We will highlight the significance of including age and gender parameters in mouse behavioral biometric-based authentication to prevent online assessment fraud. To achieve this, we will review and report key findings from several prominent and recent studies on behavioral biometric-based age and gender prediction. Given the wide range of behavioral biometric modalities discussed in the literature, our primary focus will be age and gender prediction methods based on keystroke and mouse behavioral biometrics.

2.1 Datasets and features used for age and gender prediction

i. Van Balen et al. [9] collected mouse behavior data from 94 (45 men, 49 women) participants in a controlled environment. User behavior data was collected while they performed a task-specific activity, which included identifying and clicking several targets of different sizes that change position each time it is clicked [9].
ii. Pentel [10] collected keystroke and mouse behavior data 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.

iii. Kolakowska et al. [11] also collected keystroke and mouse behavior from 42 participants (9 females and 33 males) while performing their daily browser activities. Specifically, a browser plug-in was developed to record keystroke and mouse behavior interactions. Different versions of the plug-ins were designed for different web browsers. In addition, user behavior data was collected from participants of varying age groups, including 15–24, 25–34, and 35–44.

iv. Tsimperidis et al. [12–14] acquired unconstrained keystroke behavior data to predict age and gender demographic parameters. Data was captured using a keylogger (IRecU) while participants engaged in their daily activities. According to the authors, keystroke behavior data was collected from ages 18–25, 26–35, 36–45, and 46 + age groups [12, 14, 15].

v. Tsimperidis et al. [13] acquired keystroke behavior data from 24 participants using a key-logging application. Participants typed a fixed text of 850 characters twice, once on a laptop and once on a desktop. Alongside the keystroke data, additional information regarding participants' gender and left—or right-handedness was also collected [13].

After data acquisition, raw data such as timestamp, coordinate location, action type, target location, and size are collected and used for feature extraction.

Based on the raw data, several behavioral features, including temporal, spatial, and accuracy metrics, are calculated for analysis. These metrics can be subdivided into several features, including reaction time (RT), peak velocity (PK), time to peak velocity (TPV), duration of ballistic movement (DB), the 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) [9].

On further investigation, attributes such as distance, angle, velocity, acceleration, action, and direction-based features have also been extracted and implemented for analysis [10, 11].

2.2 Classification models used for age and gender prediction

Our investigation shows that methods such as logistic regression (LR), support vector machine (SVM), decision tree (DT), *k*-nearest neighbors (*K*-NN), rotation forest (RT), Bayesian network (BNC), naïve Bayes (NB), radial basis function network (RBFN), AdaBoost (AdB), neural networks (NN), multi-layered perceptron (MLP), and random forest (RF) are popular methods for age and gender prediction [10, 11].

Many studies have been proposed to address various stages of age and gender classification, focusing on aspects such as classification methods, preprocessing techniques, and evaluation strategies. Studies implement a combination of classification methods in addition to the aforementioned methods. For example, Van Balen et al. [9] combine least-squares multiple regression (LS-MR) and LR for gender classification. In addition to the typical ML models, classifiers based on Manhattan and Euclidean distances were also implemented for classification [13].

Some research has also implemented various preprocessing methods before per- forming model training and evaluation. These include under-sampling [10, 11], normalization [11], and outlier removal [13]. On further investigation, we could also identify how several research studies evaluated their respective classification models. One of them includes using a subset of features for evaluation. For example, Van Balen et al. [9] tested several subsets of features with a certain statistical significance level for gender prediction. Similarly, Tsimperidis et al. [12, 14] also implemented information gain (IG) to determine the best feature set for age and gender prediction. Evaluation was also performed using different sets of data. For example, Van Balen et al. [9] tested the classification model using labeled and unlabeled datasets [9]. Specifically, the hold-out approach was implemented to split the dataset into train and test sets for evaluation. Tsimperidis et al. [13] also used two additional datasets to evaluate their classifiers. One was collected from a separate set of participants, while the second dataset was acquired from another source, as opposed to other studies that use a single dataset for training and evaluation [16]. In addition to the hold-out approach, several other methods, such as tenfold cross-validation [10–12, 14], have also been implemented for evaluating classifiers.

In some cases, ML models are also evaluated using different strategies [11]. For example, Kolakowska et al. [11] compared two different testing strategies, including the typical tenfold cross-validation approach and the k-fold cross-validation approach, where at each fold, one sample of each user is kept for testing. The models were also evaluated using a different number of features. For example, Tsimperidis et al. [12] evaluated various ML models

using feature sets consisting of 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500 features. This trend was also noticed in the study conducted by Tsimperidis et al. [15].

Upon further analysis, it was also observed that various features were also examined for their effectiveness in predicting age and gender [15]. For example, Tsimperidis et al. [15] conducted a study comparing the performance of keystroke duration (KD), down-down diagram latency (DDL), and their combined application.

Some studies also perform additional experimentation to understand the effect hyperparameters have on model performance. A study conducted by Tsimperidis et al. [14] developed an artificial neural network (ANN) model for age group classification and evaluated its performance by varying key parameters, including the number of hid- den layers, the number of output classes, and the learning rate and momentum values. In addition, meta-algorithms, including AdaBoost, MultiBoost, Random-correction- code, and Exhaustive-correction-code, were employed to explore ways of increasing performance rates [14, 15].

Evaluation criteria, such as accuracy (acc) [9, 12–15], f -scores [10–12, 14, 15], precision [11, 12], recall [11, 12], ROC [12, 15], and time complexity [12, 14, 15], are frequently implemented to evaluate classifier performance. Some studies have also compared several classification models to determine the best-performing model. For example, Pentel [10] compared LR, SVM, DT, and RF. Similarly, Kolakowska et al. [11] compared RT, BNC, DT, AdB, and NN. This trend was also noticed in Tsimperidis et al. [12–15].

In addition to training, testing, and evaluating classification models, a few studies have also analyzed the differences in typing speeds between age groups and genders [10]. Specifically, Pentel [10] conducted the analysis using a t-test. The study's findings indicated significant differences in typing speed across various age groups, while differences between genders were minimal. In some cases, feature analysis was also performed. For example, Kolakowska et al. [11] performed a precise analysis using the Gini index and gain ratio to determine that features like deceleration and movement speed contained the most discriminative ability. Furthermore, ANOVA analysis was performed to determine the interclass variability between groups to confirm behavioral differences [11].

2.3 Discussion and summary: age and gender prediction using behavioral biometrics

Our analysis shows that most studies analyzed in this paper collect novel data for analysis. This proves that few datasets are publicly available for mouse behavioral biometric-based age and gender prediction (Table 1). Furthermore, studies also focus on integrating several behavioral biometric modalities, such as keystroke and mouse behavioral biometrics, for age and gender prediction [10, 11].

Table 1 Studies that focus on behavioral biometric-based age and gender prediction

Author	Modality	Subjects	Dataset Type	Environment	Public
[9]	Mouse	94	Task-Specific	Controlled	No
[10]	Keystroke/Mouse	1519	Free-Form	Uncontrolled	No
[11]	Keystroke/Mouse	42	Free-Form	Uncontrolled	No
[12]	Keystroke	75	Free-Form	-	No
[13]	Keystroke	24	Free-Form	Uncontrolled	No
[14]	Keystroke	-	Free-Form	Uncontrolled	No
[15]	Keystroke	118	Free-Form	Uncontrolled	No

Compared to mouse behavioral biometrics, more studies focus on age and gender prediction using keystroke dynamics [7, 8, 12, 14, 15]. Despite this disparity, the work undertaken on age and gender prediction in different behavioral biometric modalities proves that behavior exhibited by various age groups and genders differs. In addition to the research mentioned above, research on motor behavior also indicates behavior-related differences associated with gender. Specifically, research suggests that men move faster with less accuracy than women [17, 18]. For example, [17] 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 more remarkable accuracy and slower deceleration time than men during the ballistic component of mouse movement. Similarly, [18] also studied the effect of age and gender on motor behavior. A total of 246 participants (123 males and 123 females) belonging to seven age groups were recruited for the study. Participants were required to perform a set of physical and computer tasks. Parameters such as frequency of finger tapping, 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 observed to be better in men.

The findings from the aforementioned studies highlight that age and gender models can be used as an additional layer of security to enhance the robustness of behavioral biometric online fraud detection systems. Research demonstrates that age and gender influence motor behavior, affecting movement speed, accuracy, and reaction time, which reveal distinct patterns across different

demographic groups. Given these differences, integrating age and gender characteristics can enhance the accuracy, robustness, and reliability of online assessment fraud detection systems. Further analysis also showed no publicly available datasets for mouse behavioral biometric-based age and gender prediction for our specific application.

3 Mouse behavioral biometrics for authentication: related work

We will analyze, summarize, and present findings from previously published work on mouse behavioral biometric-based authentication. Specifically, we will comprehensively analyze and report on several datasets, data collection strategies, segmentation techniques, AI methodologies, and evaluation criteria implemented by previous research studies.

3.1 Datasets and data collection strategies for mouse behavioral biometric authentication

According to our review, datasets are classified into 1) task-specific and 2) free-form datasets. Task-oriented datasets collect user behavior based on predetermined mouse operation tasks [7, 9]. Meanwhile, free-form datasets collect mouse behavior data by continuously monitoring users' daily activities without specific instructions [7, 9]. Based on the data collected, either static or dynamic authentication is performed.

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 authentication.

3.1.1 Novel datasets collected for mouse behavioral biometric authentication

i. Subash et al. [8] collected novel mouse behavior data while participants engaged in an online education game. Their methodology included collecting data from 13 participants who were required to perform three assessment-like tasks, including an MCQ, click-the-target, and matching tasks. In addition to collecting mouse behavior, participants were also required to fill out a pre-participation questionnaire containing questions related to the participants' demography and computer proficiency [8].
ii. Zheng et al. [7] collected two datasets: 1) a controlled and 2) an uncontrolled set. The first dataset was collected from 30 participants in a controlled environment. According to the authors, the participants were from diverse age groups, held various occupations, and possessed different levels of educational qualifications. Their methodology required participants to perform routine activities, including surfing the Internet, programming, online chatting, and gaming. The second dataset collected mouse behavior data from 1000 participants in an uncontrolled manner [7].
iii. Siddiqui et al. [19] collected novel mouse behavior data from 10 participants in a controlled environment. Ten participants were recruited and required to play a game of Minecraft for 20 min. Their methodology required participants to play the game using the same device (desktop) and was subjected to the same default game settings. A Python program was implemented to record the mouse behavior during the session.
iv. Wang et al. [20] collected mouse behavior data from 18 participants. Their methodology required subjects to perform two tasks after their emotions were aroused. Several videos are used for this purpose. Specifically, three videos stimulate positive, negative, and neutral emotions. Furthermore, a face reader is also used to detect emotional changes. A well-structured academic website is developed for data collection. Participants were required to perform two tasks immediately after they watched the video [20].
v. Shen et al. [21] collected novel mouse behavior data from 37 participants in a controlled environment. Their methodology required participants to perform two rounds of data collection per day and keep a 24-h gap between the collections. The data collection procedure involves performing a task-specific mouse activity ten times each. Specifically, the task involved clicking several targets (buttons) prompted by the application. Every two adjacent movements were separated by either a single or double click. The task comprises 16 mouse moves, eight single-click events, and eight double-click events. According to the author, the participants were requested to use only the external mouse device while they performed the activity.
vi. Da Silva et al. [22] collected both keystroke and mouse behavior data from 55 participants in a controlled manner. Their methodology required five participants to play a League of Legends game, which lasted between 30 and 50 min. According to the author, each participant could choose which computer and character to play with. A background application was developed using C# for data acquisition.
vii. Feher et al. [17] collected mouse behavior data from 25 subjects from different groups: 1) Internal and 2) External subjects. According to the author,

the systems used for data collection were chosen from various brands and hardware configurations. Furthermore, one or more internal subjects are authorized to interact with a particular system, while the rest are not.

3.1.2 Public datasets for mouse behavioral biometric-based authentication

i. Balabit Dataset: This is a publicly available dataset that collected mouse behavior data from 10 users in an uncontrolled manner. Data was collected from the participants while they were working over remote desktop clients connected to remote servers. The data is divided into training and test files, each stored separately [14, 23, 24].
ii. DFL Dataset: This publicly available dataset collected mouse behavior data from 21 participants in an uncontrolled environment. User behavior data was collected while participants performed routine work on their systems. Hence, data was collected from several devices, including desktops and laptops. Furthermore, data also captured the user behavior interaction performed on different input hardware devices, such as external mouse and trackpads [23].
iii. Choa Shen Dataset: This publicly available dataset collected mouse behavior data from 28 participants while they performed their routine work. Like the DFL dataset, the data acquisition was performed using a background recording service [23].

3.2 Pre-processing methods and mouse behavioral features for authentication

After data collection, basic raw data, such as screen coordinates (Crd) [7, 8, 25], timestamp (t), [7, 8, 25], action type (AT), [7, 8, 25], screen height and width (SH, SW) [7, 8, 25] are used for feature extraction. Features such as horizontal velocity (HV), vertical velocity (VV), acceleration (Acc), jerk (j), angular velocity (AV), and curvature (C) are extracted [8, 19, 25]. However, these features cannot be sent for analysis as they are calculated based on individual mouse events, which cannot comprehensively profile user behavior. Therefore, a pre-processing method called segmentation is implemented for mouse behavioral biometric-based authentication.

Segmentation is a process that divides mouse behavior data into meaningful and logical blocks of information. It is implemented to acquire aggregate features that help in profiling users for effective user authentication. Our investigation shows that different segmentation methodologies, including point-and-click (PC), drag-and-drop (DD), mouse movement (MM), pause-and-click (PaC), and image-based segmentation techniques, have been implemented for mouse behavioral biometric-based authentication [7, 8, 25].

After segmentation, several aggregate mouse behavior features, such as average (Avg), standard deviation (std), minimum (min), maximum (max), and range (range) of the features mentioned above, are extracted and implemented for analysis. Our investigation identified over 70 mouse behavior features used by previously published work (Fig. 1). For easy understanding, we represent all mouse behavior features pictorially by grouping them based on popularity. Specifically, we first identify the most popular features (features used in more than 50% of studies), followed by features used by 36–50% of studies, 26–36%, and the least popular features.

In addition to those mentioned above, additional features such as sum of angle (SOA), number of points (NOP), number of critical points (NOC), straightness (SR), trajectory length (TL), acceleration at the beginning (ABT), sharp angles (SA), most significant deviation (LD), type of action (TA), and jitter have also been extracted and used for analysis.

3.3 AI methodologies and evaluation criteria implemented for mouse behavioral biometric authentication

Based on our investigation, mouse behavioral biometric-based authentication relies on several ML and deep learning (DL) approaches. Approaches like SVM [7, 20], RF [19, 25], and *K*-NN [20] are commonly used for analysis.

In addition to typical ML approaches, studies also implement DL approaches, such as convolutional neural networks (CNN) [22], multi-layered perceptron (MLP) [20], and recurrent neural networks (RNN-LSTM). Compared to typical conventional ML models that rely on statistical features for classification, most DL models analyzed use a sequence of raw mouse behavior data for analysis [22, 24]. In some studies, DL models were evaluated using conventional statistical features. For example, Subash et al. [8] assessed the RNN-LSTM model with a unique set of traditional features. In certain instances, DL models were also evaluated using images derived from recorded mouse behavior [24]. Hu et al. [24] conducted a study in which mouse behavior data were represented as images for analysis. Specifically, the mouse movement data was mapped onto coordinate space to generate visual representations, which were then analyzed using a CNN. Table 3 summarizes the various models used for mouse behavioral biometric authentication.

Some studies also implement various pre-processing methods, in addition to segmentation approaches, before

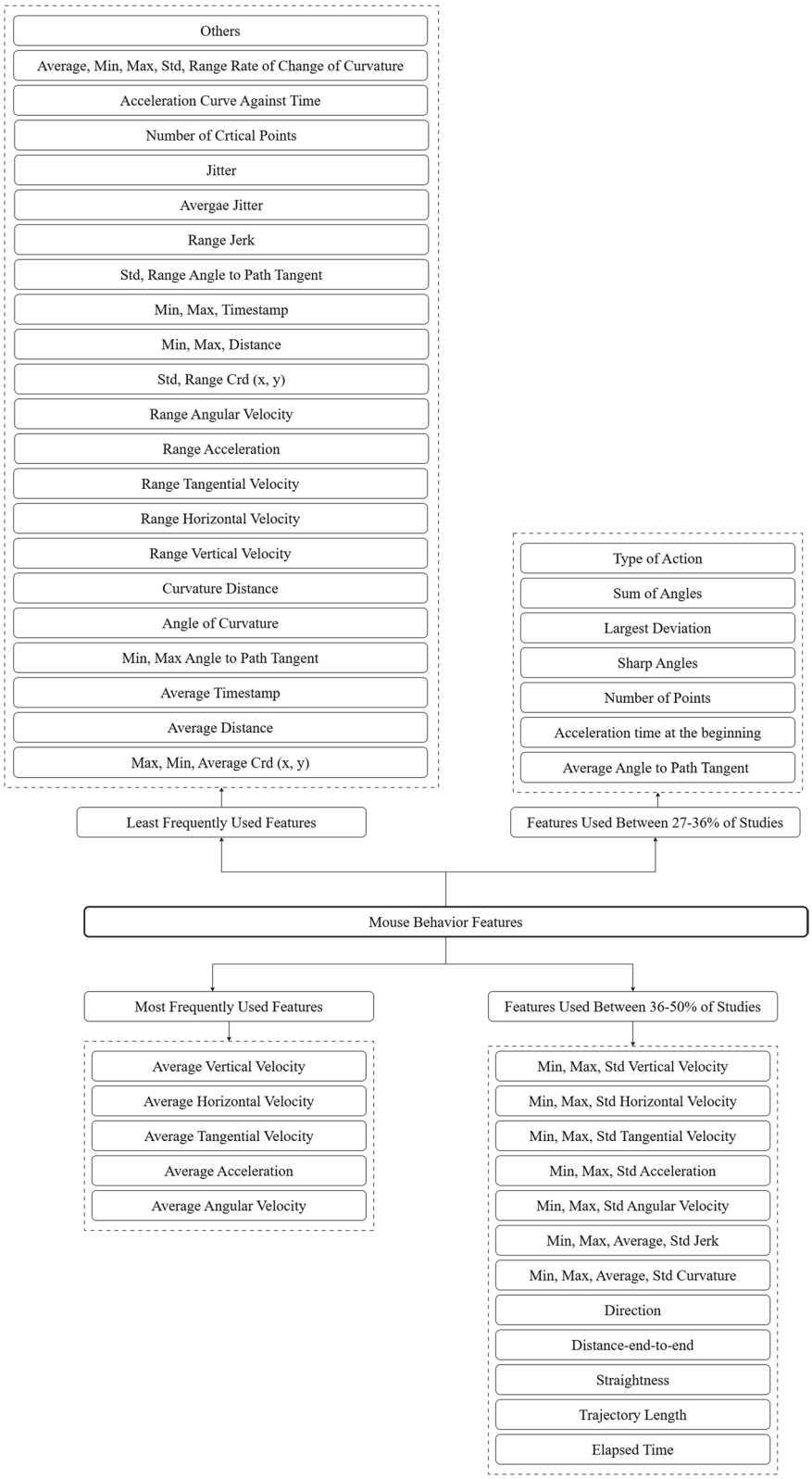

Fig. 1 Mouse behavior features are organized from most popular to least popular

training and evaluating classifiers. These include normalization [8], SMOTE oversampling techniques [8], signal smoothening [25], and bootstrapping [21]. In addition to this, we were also able to identify how several classification models were evaluated. Some studies compared several features to determine the best feature set for user authentication. For example, Subash et al. [8] compared several features, including the commonly implemented features gathered from previously published work, to identify a unique set of features known as the user behavior-centric feature set for robust mouse behavioral biometric-based authentication.

Studies have also identified a unique set of features relatively independent of the operating environment. For example, Zheng et al. [7] identified angle-based features, including direction (Dir), angle of curvature (AOC), and curvature distance (CD). Similarly, Shen et al. [21] also evaluated a unique set of procedural and holistic features for user authentication. In addition, the study was also able to partially mitigate behavior variability by implementing distance metrics and kernel PCA to obtain distance-based eigenspace to represent mouse behavior feature space. Some research studies also use different evaluation scenarios to evaluate their proposed classifiers. For example, Zheng et al. [7] evaluated their proposed SVM classifier using datasets collected from different environments. Specifically, the classifier was trained using data collected from a desktop in the work environment and tested using a dataset collected from a laptop in the home environment.

In addition to this, the classifier's performance was compared with different numbers of PC actions, ranging from 1 to 25. This study also compares PC and partial-movement (PM) segmentation approaches. According to the study, including the PM segmentation approach in the analysis degrades classifier performance. Similarly, Siddiqui et al. [19] evaluated their RF classifier under two distinct scenarios. In the first scenario, the classifier was trained and tested using the same dataset. In the second scenario, the classifier was trained using a training set and evaluated on a separate testing set. A similar approach was observed in the study by Antal and Egyed-Zsigmond [25]. Furthermore, their classifier was assessed using the MM, PC, and DD segmentation approaches individually and in a combined configuration [25]. This same trend was noticed in a study by Shen et al. [21]. In addition to this, classifiers were tested using features extracted from different sample lengths. This was also noticed in a study performed by Hu et al. [24]. Studies also compare several classifiers for evaluation. For example, a study conducted by Da Silva et al. [22] compares *K*-NN, SVM, MLP, and RF. Similarly, Subash et al. [8] compare RNN-LSTM, MLP, SVM, NB, *K*-NN, RF, and DT. This trend is also observed in Shen et al. [21]. Based on further investigation, we find that some studies compare classifier performance using different datasets. For example, Antal and Denes-Fazakas [23] tested the classifier using three publicly available datasets: Balabit, DFL, and Choa Shen. Furthermore, a comprehensive comparative analysis was performed by comparing the performance of the classifier using different numbers of actions [23]. Evaluation criteria such as accuracy (acc), precision (pre), recall (rec), false rejection rates (FRR), false acceptance rates (FAR), authentication time (AT), area under the curve (AUC), and equal error rate (EER) have commonly been used to assess classifier performance [7, 8, 19–25].

3.4 Discussion and summary of mouse behavioral biometric authentication research

Based on our background analysis (Table 2), most of the research analyzed in the study focuses on collecting data for mouse behavioral-biometric-based authentication. Further investigation revealed little research on mouse behavioral-biometric-based assessment fraud detection in online education platforms. Furthermore, there are no publicly available datasets for this specific application (Table 2).

Among the novel datasets, free-form datasets are frequently employed for analysis. However, despite the availability of publicly accessible free-form datasets, a subset of studies continues to utilize task-specific datasets, driven by the requirement for more specialized data tailored to their analysis needs (Fig. 2). Some studies also merge several modalities for mouse behavioral biometric-based authentication. For example, Da Silva et al. [22] combined keystroke and mouse behavior features for enhanced behavioral biometric-based authentication. According to the study, combining keystroke and mouse behavior features resulted in increased performance rates, achieving the highest accuracy of 90% using RF. Similarly, this trend is also noticed in Traore et al. [26], Panasiuk et al. [16], and Mondal & Patrick Bours [27]. Further investigation found that most studies still rely on conventional ML algorithms despite the availability of advanced DL approaches. Furthermore, we identify the RF algorithm as the most popular ML model implemented for mouse behavioral biometric authentication. Our analysis also revealed that most mouse behavioral biometric-based authentication models suffer from high false positive rates [8, 16–22, 24, 26, 27].

The extensive background analysis indicates that combined modality models, particularly those integrating keystroke dynamics and mouse behavior, demonstrate superior performance compared to individual modality models. However, implementing such models is often

Table 2 Summary of datasets and data collection strategies used in mouse behavioral biometric-based authentication

Author	Dataset Name	Public	Subjects	Environment	Dataset Type
[8]	Novel Dataset	No	13	-	Task-Specific
[7]	Novel Dataset	No	30	Controlled	Free-Form
			1000	Uncontrolled	
[19]	Novel Dataset	Yes	10	Controlled	Free-Form
[25]	Balabit Dataset	Yes	10	Uncontrolled	Free-Form
[20]	Novel Dataset	No	18	Controlled	Task-Specific
[21]	Novel Dataset	Yes	37	Controlled	Task-Specific
[22]	Novel Dataset	No	55	Controlled	Free-Form
[23]	Balabit Dataset	Yes	10	Uncontrolled	Free-Form
	Choa Shen Dataset	Yes	28	Uncontrolled	
	DFL Dataset	Yes	21	Uncontrolled	
[24]	Balabit Dataset	Yes	10	Uncontrolled	Free-Form
[17]	Novel Dataset	No	25	Controlled	-
[18]	Balabit Dataset	Yes	10	Uncontrolled	Free-Form
	DFL Dataset	Yes	21	Uncontrolled	

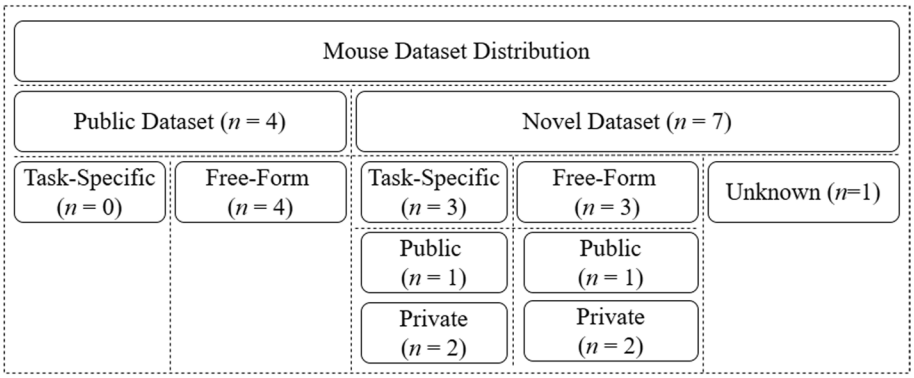

Fig. 2 Mouse behavioral dataset types, distribution, and availability

complex and cumbersome. Furthermore, the analysis presented in previous sections has indicated that integrating age and gender demographic parameters can be used as an additional layer of security to enhance the robustness and classification rates of mouse behavioral biometric online fraud detection systems. Therefore, to reduce the false positive rates currently identified in studies, we propose a UserID.AGE.GEN framework (Sect. 6.1) implements a novel cross-referencing fusion algorithm that integrates age and gender demographic parameters for enhanced user authentication.

4 Data collection strategy

Our main objective is to integrate age and gender demographic parameters in mouse behavioral biometric-based authentication for online fraud detection in online education platforms. Hence, we need to collect both mouse behavior and demographic information.

Our background analysis shows no publicly available datasets for our specific application (Table 3). Therefore, an online assessment game containing four assessment-like tasks was developed to collect mouse behavior data. These include click-the-target, MCQ, drag-drop, and matching tasks. The rationale for incorporating multiple assessment-like tasks into the data collection procedure was to comprehensively cover online assessment types, input styles, and behavior information. This enabled us to diversify the data collection and collect sufficient mouse behavior data for effective mouse behavioral biometric-based online assessment fraud detection. The description of tasks is as follows:

i. MCQ Task: Contains four simple general knowledge questions that participants answer by selecting the correct choice among four given options. The subsequent question is displayed only when

Table 3 AI methods implemented for mouse behavioral biometric authentication

Author	Machine Learning Approaches	Deep Learning Approaches
[8]	-	RNN-LSTM
[7]	SVM-RBF	-
[19]	RF	-
[25]	RF	-
[20]	*K*-NN, SVM, RF	MLP
[21]	RF	-
[22]	-	CNN
[23]	RF	-
[24]	-	CNN
[17]	RF	-
[18]	SVM	-

the current question is answered correctly. Once a choice is selected, the participant must click the submit button. If the option 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.

ii. 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, which includes mammal, amphibian, reptile, fish, and bird. In total, there are five images and five drop-boxes. If the participant drags the image into 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 a participant places an image into an incorrect dropbox, the image will automatically return to its original position, and the background will briefly change to red before reverting to its initial state. Participants are allowed to correct their mistakes.

iii. 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 acquisition was done with the help of a web application developed using HTML, CSS, and JavaScript. Specifically, mouse event listeners were implemented to record the performed mouse action. Data was collected from 20 participants recruited from Sanjay Gandhi College of Education, Bangalore, India. All participants were required to complete four types of tasks—click-the-target (CT), MCQs, drag-drop (DD), and matching (MA) tasks during each data collection session. MCQ, DD, and MA tasks were performed ten times per session, resulting in a structured and repetitive interaction sequence designed to capture consistent behavioral biometric patterns. Data collection was conducted over one year, with sessions spaced at monthly intervals to allow for the observation of temporal variations in user behavior. Participants were instructed to access the website using a desktop or a laptop computer. Participants were advised to use the university computer laboratories to complete the tasks when they did not have access to a personal device.

In addition to mouse behavior data, we collected user demographic (age group and gender) information via a pre-participation questionnaire. Among the 20 participants, 10 are female, and 10 are male, distributed across two age groups: 18- 22 and 23–27. During task engagement, raw data, such as timestamp, screen height, screen width of the content area, coordinates (X, Y), action types, element on which the event was performed, offsetX, and Y, are collected for further feature extraction. Mouse behavior data is received individually for each user and task type in JSON format. We have collected around 41,400 rows of raw mouse behavior data for our experimentation. It is important to note that the proposed approach will be evaluated using the first trial from the first session.

5 Feature engineering

Before performing model training or testing, we perform segmentation, a critical pre-processing step that groups raw mouse behavior data into logical blocks of information [7, 8, 19, 25]. After segmentation, aggregate features such as mean, standard deviation, minimum, maximum, and range are extracted for analysis. This study compares three segmentation methods, namely 3, 5, and 10 mouse-event (MM) segmentation methods. The use of various segmentation methods in prior research (see Sect. 3.2) underscores the absence of a standardized approach for segmentation in this domain. Some studies utilize image-based segmentation techniques, others adopt point-and-click (PC) methods, while others rely on MM events to guide the segmentation process. Given this variability, we adopt the comparative framework established in our previous work [8], wherein segmentation strategies based on 3MM, 5MM, and 10MM events are evaluated.

These segmentation methods fall the under the category of nMM segmentation, where n refers to the number of mouse events that need to be considered to form a logical block (segment) of mouse behavior data. Using

these segmentation methods, we extract the following aggregate features to represent user profiles (Table 4). In total, 40 features were implemented for analysis. Specifically, these features have been implemented for age, gender, and user ID prediction.

6 Experimentation

Before integrating age and gender parameters to bolster mouse behavioural biometric-based assessment fraud detection, we will perform age, gender, and user ID prediction separately to observe their performance.

To perform prediction analysis, we follow the methodology illustrated in Fig. 3. We perform segmentation, feature extraction, model training, and evaluation. Specifically, we will compare several segmentation methods, feature sets, and ML algorithms to determine the most suitable combination for age, gender, and user ID prediction.

Like we previously mentioned, given the variability of segmentation methods, we adopt the comparative framework established in our previous work [8], wherein segmentation strategies based on 3MM, 5MM, and10MM events are evaluated. This experimentation enables us to determine the most effective combination of ML algorithm and segmentation method to be implemented in our age, gender, and user ID prediction models.

Once we identify the best segmentation and ML approach, we apply two feature selection (FS) techniques to the original feature set to determine the most effective subset of features for our proposed approach. Specifically, we compare the feature sets obtained through Recursive Feature Elimination (RFE) and XGBoost-based feature importance. This step is motivated by the variety of features explored in previous studies (3.2), which makes it cumbersome to manually determine the most relevant ones. Employing these feature selection methods facilitates a more systematic and data-driven identification of the optimal feature set.

Based on the experimental evaluations described above, we will identify the best combination of segmentation method, feature set, and ML approach for our age, gender, and user ID prediction models.

Table 4 Original features: features extracted for age, gender, and user ID prediction

Features	Aggregated Features	Dimensions
X, Y Crd	Mean, Min, Max	6
Distance	Mean, Min, Max	3
Tangential Velocity	Mean, Min, Max	3
Horizontal Velocity	Mean, Min, Max	3
Vertical Velocity	Mean, Min, Max	3
Acceleration	Mean, Min, Max	3
Angular Movement	Mean, Min, Max	3
Angular Velocity	Mean, Min, Max	3
Timestamp	Mean, Min, Max	3
Jerk	Mean, Min, Max	3
Curvature	Mean, Min, Max	3
Sum of Angles	Stand-alone Feature	1
Distance End-to-End	Stand-alone Feature	1
Trajectory Length	Stand-alone Feature	1
Elapsed Time	Stand-alone Feature	1
Total Features		40

7 Results

Tables 5 and 6 present the experimental results corresponding to the methodology outlined in 3. These tables report the predictive performance for age, gender, and user ID across various ML algorithms and segmentation

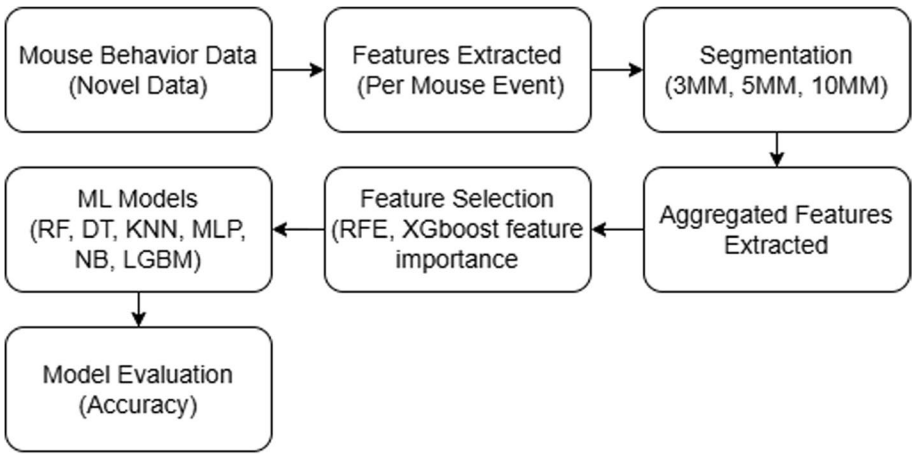

Fig. 3 Methodology for age and gender prediction

Table 5 Comparison of segmentation methods for age and gender prediction

Segmentation	Algorithm	Accuracy Age Prediction	Accuracy Gender Prediction
3MM	RF	79.98%	78.82%
	KNN	64.38%	61.68%
	MLP	67.25%	62.79%
	NB	57.83%	54.52%
	LGBM*	87.69%	83.96%
5MM	RF	74.97%	76.60%
	KNN	63.73%	60.77%
	MLP	61.18%	62.81%
	NB	58.52%	56.07%
	LGBM	83.14%	83.04%
10MM	RF	74.18%	77.04%
	KNN	57,58%	60.04%
	MLP	61.47%	53.27%
	NB	56.35%	51.02%
	LGBM	84.42%	86.47%

*indicates the highest performing accuracy for the respective feature set or algorithm

Table 6 Comparison of segmentation methods for user ID prediction using original features

Segmentation	Algorithm	Accuracy for User ID Prediction
3MM	RF	67.19%
	KNN	22.88%
	MLP	29.55%
	NB	13.70%
	LGBM*	75.15%
5MM	RF	60.87%
	KNN	19.71%
	MLP	23.39%
	NB	14.30%
	LGBM	71.50%
10MM	RF	56.35%
	KNN	19.46%
	MLP	25.61%
	NB	14.34%
	LGBM	72.95%

*indicates the highest performing accuracy for the respective feature set or algorithm

methods, evaluated using the original feature set in Table 4. The results indicate that the feature set in Table 4 is well-suited for predicting age, gender, and user ID. Moreover, combining the 3MM segmentation method with the LGBM algorithm consistently yields the highest performance across evaluation metrics, outperforming other algorithmic and segmentation configurations. This observation aligns with the findings reported in our earlier studies [8, 28]. We have included Table 9 in the appendix, which explains the hyperparameters used in the ML algorithms. The 3MM segmentation method was selected because it yielded superior performance for both the user identification and age prediction models, despite the gender prediction model demonstrating better results with the 10MM segmentation method.

The LGBM model was also observed to consistently outperform all other ML models, regardless of the segmentation method employed. Using LGBM, we achieve satisfactory performance of 87.69%, 83.96%, and 75.15% accuracy for age, gender, and user ID prediction, respectively.

7.1 Evaluating different feature sets for mouse behavioural biometric age, gender, and user ID prediction

Based on the experimentation in the previous section, we determined that the LGBM ML approach paired with the 3MM segmentation method outperforms other approaches tested, achieving an accuracy of 87.69%, 83.96%, and 75.15% for age, gender, and user ID prediction, respectively.

In this section, we will compare three different feature sets, including the original set with a subset of features resulting from the RFE and XGBoost feature importance FS methods. We implement FS mainly to reduce the dimensionality of the input features to the model and to automate the feature extraction process, as the original feature set was manually identified based on previous literature. The main objective of this section is to determine the best feature set to integrate with the RF and 3MM segmentation method. To identify the best feature sets, we compare three feature sets for age, gender, and user ID prediction. Table 7 illustrates the evaluation of several features.

Based on the results presented in Table 7, the feature set derived from the XGBoost feature importance method demonstrates the highest compatibility with the 3MM segmentation method when combined with the LGBM algorithm.

The XGBoost-selected feature set was adopted as it led to a ~ 1–2% improvement in the performance of both the gender classification and user identification models compared to the original feature set. Notably, for age prediction, the performance using the XGBoost features was comparable to that achieved with the original features, with only a marginal difference in accuracy. Therefore, the XGBoost-selected features were chosen for subsequent analysis. This feature set includes *max_curvature*, *elapsed_time*, *max_timestamp*, *max_jerk*, *min_timestamp*, *min_distance*,

Table 7 Comparison of several feature sets for age, gender, and user ID prediction

Feature Set	Accuracy (%)	Application	Segmentation	ML Algorithm
Original Feature Set	87.69	Age Prediction	3MM	LGBM
RFE Features	82.80			
Xgboost Features*	87.51			
Original Feature Set	83.96	Gender Prediction	3MM	LGBM
RFE Features	83.84			
XGboost Features*	85.49			
Original Feature Set	75.15	User ID Prediction	3MM	LGBM
RFE Features	70.56			
XGboost Features*	75.76			

*indicates the highest performing accuracy for the respective feature set or algorithm

min_vertical_velocity, *mean_timestamp*, *max_Acceleration*, *min_x*, *max_y*, *max_x*, *max_distance*, *max_Angular_Velocity*, and *min_y* (A total of 15 features).

The results obtained using the XGBoost feature set are 87.51%, 85.49%, and 75.76% for predicting age, gender, and user ID, respectively. This feature set demonstrates improved performance relative to the original feature set (Table 4), which achieved 87.69%, 83.96%, and 75.15% for the same prediction tasks.

7.2 Mouse behavioral biometric-based online assessment fraud detection using age and gender parameters

The results presented in the previous section indicate that combining the LGBM algorithm with the 3MM segmentation method and the XGBoost-derived feature set yields the most effective configuration for predicting user ID, age, and gender. This approach achieved prediction accuracies exceeding 80% for age and gender classification, and over 75% for user ID prediction.

In this section, we will improve the performance of the user-id prediction model by proposing a novel USERID.AGE.GEN framework (Fig. 4) for mouse behavioral

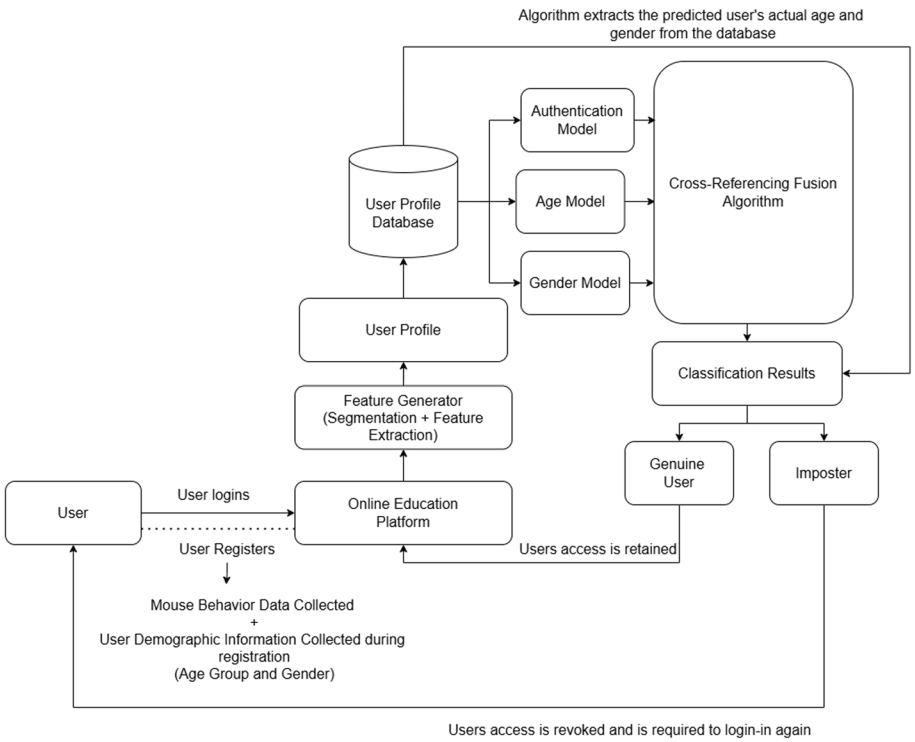

Fig. 4 USERID.AGE.GEN framework for authentication for online education platforms

biometric-based authentication for online education platforms for online assessment fraud detection. This framework implements a novel cross-referencing algorithm (CRFA) integrating age and gender parameters to bolster performance. Specifically, this algorithm combines the results obtained from the independent mouse behavioral biometric-based user ID, age, and gender models (Fig. 4). We will use a specifically designed algorithm to accomplish our proposed approach.

This algorithm (Fig. 5) is designed to integrate the age and gender models into the decision-making process of an ML or DL-based online assessment fraud detection system. Our primary assumption is that the user's age group and gender are previously known. The CRFA algorithm has been represented using a flowchart (Fig. 5). To combine the decision of the three classifiers, we calculate the fusion probability for all possible combinations of age, gender, and user ID labels by multiplying each probability value present in the prediction probabilities arrays obtained from evaluating the User ID, age, and gender models (Fig. 5). The formula used for calculating fusion probability is mentioned in Fig. 5.

The argmax() function is then implemented to extract the highest probability product from the set of combinations we calculated earlier. The predicted user ID, age, and gender are extracted from the highest fusion probability value. The predicted user ID is also used to get the actual age and gender of the predicted user ID, as we already know the actual age and gender of the user. Comparison is performed between the predicted age/gender and the actual age/gender of the predicted user ID. We assume that if the predicted age/gender matches the actual age/gender, the User ID model's prediction is correct. Otherwise, the classification is wrong. If the classification is incorrect, we sort the user ID prediction probability array and extract the second-highest fusion probability value from the set of combinations we calculated earlier.

Based on the second-highest fusion probability, the predicted user ID, age, and gender are extracted. The same check between the predicted age/gender and the actual age/gender values is performed. This step is iterative until a correct classification is found. This additional step in the flowchart allows the model to iteratively refine its predictions by systemically re-evaluating and repeating the classification process based on the following highest probability values, leading to more reliable results. Due to this particular enhancement, we name our approach an enhanced cross-referencing algorithm (ECRA).

Based on the results (Table 8), we can conclude that our proposed method outperforms our stand-alone user ID model methods by ~ 1%, achieving 76%, 75%, and 73% accuracy, precision, and recall. One notable limitation of this paper is that the model's performance can be further improved by considering other user-centric parameters.

An additional experiment was conducted using the original feature set to evaluate the robustness of the ECRA model. Specifically, the ECRA model was evaluated and demonstrated a performance improvement of approximately ~ 1–2% over the LGBM algorithm. The LGBM model achieved accuracy, precision, and recall scores of 75.15%, 73.71%, and 72.49%, respectively, whereas the ECRA model attained enhanced scores of 76%, 75%, and 74%. Notably, this performance gain is more noticeable than the improvements observed when using the XGBoost-selected feature set, further highlighting the effectiveness of the ECRA model.

Based on further analysis, we confirm that our proposed model demonstrates improved performance compared to certain previously published studies, such as [19] and [16], the latter utilizes a fusion model combining keystroke and mouse behavior features and reports an accuracy of 68.80%. However, it is important to acknowledge that a direct comparison is limited because these studies were evaluated using different datasets and evaluation criteria, which may vary in characteristics and complexity [7, 8, 16–26].

8 Privacy and ethical considerations in behavioral biometric authentication

The integration of demographic parameters such as age and gender in user identity verification systems for online education platforms must be ethically grounded in principles of fairness, necessity, and proportionality. The primary justification lies in enhancing the accuracy and reliability of identity verification mechanisms, which are critical in maintaining the integrity of online assessments and access control. Age and gender, as behavioral biometric correlates, can contribute meaningfully to the discrimination of users in continuous authentication frameworks by leveraging subtle variations in interaction patterns. However, the ethical application of such demographic attributes must ensure that their use does not reinforce bias, lead to discrimination, or compromise user privacy. To this end, datasets must be handled with strict adherence to data protection regulations (e.g., GDPR), ensuring that demographic data is anonymized, securely stored, and used solely for authentication. Furthermore, any deployment of such systems should be accompanied by transparent user consent processes and options for opting out.

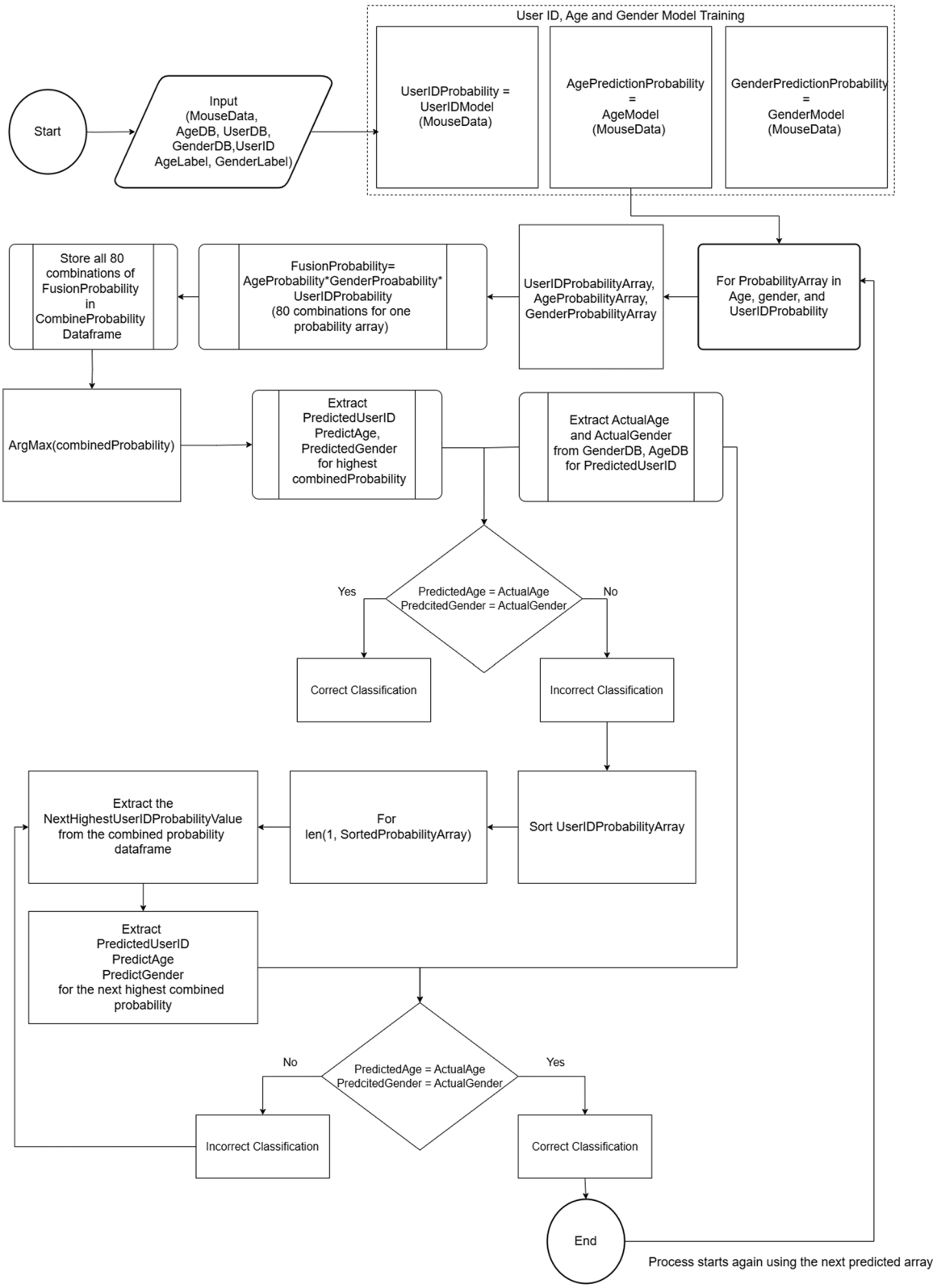

Fig. 5 Flowchart illustrating enhanced cross-referencing algorithm (CRFA)

Table 8 Comparison of our proposed ECRA approach and stand-alone user ID model

Model	ML Algorithm	Accuracy(%)	Precision(%)	Recall(%)
Stand-Alone User ID Model	LGBM	75.76	74.60	72.95
ECRA	LGBM	76%	75%	73%

In summary, while the use of age and gender in behavioral biometric-based authentication can be ethically justified by its functional contribution to identity verification and system integrity, it must be balanced by rigorous safeguards that protect individual rights and prevent misuse or bias.

9 Conclusion

In this paper, we present our new USERID.AGE.GEN framework implements a new cross-referencing algorithm capable of integrating age and gender models to enhance the detection of behavioral biometric-based mouse assessment fraud for online education platforms. To accomplish this, we collected novel data while participants played an online education game consisting of multiple online assessment-like tasks. In addition to mouse behavior data, demographic parameters such as age group and gender were collected and implemented for analysis. Our extensive experimentation yielded a performance greater than 85% accuracy for age group and gender prediction. From this experimentation, we were able to determine the implementation of age-group and gender prediction models to improve the performance rate of the mouse behavioral biometric-based online assessment fraud detection. Based on our experimentation, we confirmed that our proposed model could increase the performance of the user authentication model by 1–2%, thereby demonstrating its capability to improve performance.

Overall, the findings presented in this paper support further research into integrating different user-centric attributes, similar to user demographic parameters, to improve online assessment fraud detection. In our future works, we aim to collect additional and more diverse data across multiple sessions, including participants from a broader range of age groups and educational institutions. This will enable a more comprehensive evaluation of the proposed model, particularly its generalizability to larger datasets and robustness to evolving user behavior patterns over time. Furthermore, we plan to conduct validation studies in operational online education environments, involving diverse user populations, varied interaction contexts, and naturally occurring behaviors to understand model performance under realistic conditions. This evaluation will also provide insights into deployment challenges like scalability and user acceptance. Furthermore, we plan to extend the experimentation to include state-of-the-art DL approaches like CNN, RNN, Transformers, and their hybrid variations. This will help refine our approach, ensuring adaptability to evolving user patterns and enhancing the model's reliability, robustness, and generalizability in dynamic environments.

Appendix

Machine learning hyperparameters values

We have added Table 9, explaining the hyperparameters used in the ML algorithms.

Hyperparameters used in ML algorithms

ML Algorithms	Hyperparameters Used
RF	n_estimators = 1000, max_depth = None, min_samples_split = 2, min_samples_leaf = 1, class_weight = 'balanced', random_state = 42
*K*NN	n_neighbors = 6, weights = 'uniform', algorithm = 'auto', n_jobs = -1
MLP	hidden_layer_sizes = (60, 64, 32, 25), activation = 'relu', solver = 'adam', learning_rate_init = 0.001, max_iter = 1000 random_state = 42
NB	var_smoothing = 1e-9
LGBM	n_estimators = 1000

Link to public datasets

1. Dataset collected by Siddiqui et al., 2021 [19]: https://github.com/NyleSiddiqui/MinecraftMouse-Dynamics-Dataset
2. Dataset Collected by Shen et al., 2012 [21]: http://nskeylab.xjtu.edu.cn/projects/mousedynamics/behavior-data-set/
3. Balabit Dataset: https://github.com/balabit/Mouse-Dynamics-Challenge
4. DFL Dataset: http://www.ms.sapientia.ro/~manyi/DFL.html
5. Chao Shen Dataset: http://nskeylab.xjtu.edu.cn/projects/mousedynamics/monitoring/

Acknowledgements
NA

Authors' contributions
A.S: Conceptualization, validation, investigation, data collection, coding, methodology, visualization, writing and editing. I.S: Conceptualization, supervision, validation, writing, methodology, review. I.L: Supervision, validation, review. K.L: Supervision, validation, review.

Funding
NA.

Data availability
The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
All data has been collected by following ethical guidelines, which include the approval of ethics applications and collection of participant consent.

Competing interests
The authors declare no competing interests.

Received: 7 March 2025 Accepted: 26 June 2025
Published online: 03 July 2025

References

1. Z.M. Basar, A.N. Mansor, K.A. Jamaludin, B.S. Alias, The effectiveness and challenges of online learning for secondary school students–A case study. Asian Journal of University Education **17**(3), 119–129 (2021)
2. V. Mili´cevi´c, N. Deni´c, Z. Mili´cevi´c, L. Arsi´c, M. Spasi´c-Stojkovi´c, D. Petkovi´c, J. Stojanovi´c, M. Krkic, N.S. Milovan˘cevi´c, A. Jovanovi´c, E-learning perspectives in higher education institutions. Technological Forecasting and Social Change **166**, 120618 (2021)
3. D. Portugal, J.N. Faria, M. Belk, P. Martins, A. Constantinides, A. Pietron, A. Pitsillides, N. Avouris, C.A. Fidas, Continuous user identification in distance learning: a recent technology perspective. Smart Learning Environments **10**(1), 38 (2023)
4. M. Garg, A. Goel, A systematic literature review on online assessment security: Current challenges and integrity strategies. Computers Security **113**, 102544 (2022)
5. E. Mohammadkarimi, Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. Journal of Applied Learning and Teaching **6**(2), 105–113 (2023)
6. F. Noorbehbahani, A. Mohammadi, M. Aminazadeh, A systematic review of research on cheating in online exams from 2010 to 2021. Educ. Inf. Technol. **27**(6), 8413–8460 (2022). https://doi.org/10.1007/s10639-022-10927-7
7. N. Zheng, A. Paloski, H. Wang,. An efficient user verification system via mouse movements. In Proceedings of the 18th ACM conference on Computer and communications security (USA. ACM H, New York NY 2011) p. 139–150
8. A. Subash, I. Song, I. Lee, K. Lee, Mouse Dynamics-Based Online Fraud Detection System for Online Education Platforms. In: XS. Yang, S. Sherratt, N. Dey, A. Joshi, (eds) Proceedings of Ninth International Congress on Information and Communication Technology. ICICT 2024 2024. Lecture Notes in Networks and Systems vol 1003, (Springer, Singapore, 2024)
9. N. Van Balen, C. Ball, H. Wang, Analysis of Targeted Mouse Movements for Gender Classification. EAI Endorsed Transactions on Security and Safety **4**(11), 153395–153415 (2017). https://doi.org/10.4108/eai.7-12-2017.153395
10. A. Pentel, Predicting age and gender by keystroke dynamics and mouse patterns. In Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (USA ACM, New York NY, 2017), pp. 381–385
11. A. Kolakowska, A. Landowska, P. Jarmolkowicz, M. Jarmolkowicz, K. Sobota, Automatic recognition of males and females among web browser users based on behavioural patterns of peripherals usage. Internet Res. **26**(5), 1093–1111 (2016)
12. I. Tsimperidis, A. Arampatzis, A. Karakos, Keystroke dynamics features for gender recognition. Digit. Investig. **24**, 4–10 (2018)
13. I. Tsimperidis, V. Katos, N. Clarke, Language-independent gender identification through keystroke analysis. Information and computer security **23**(3), 286–301 (2015)
14. I. Tsimperidis, S. Rostami, V. Katos, Age detection through keystroke dynamics from user authentication failures. International journal of digital crime and forensics **9**(1), 1–16 (2017)
15. I. Tsimperidis, C. Yucel, V. Katos, Age and gender as cyber attribution features in keystroke dynamic-based user classification processes. Elec tronics (Basel) **10**(7), 835 (2021)
16. P. Panasiuk, M. Szymkowski, M. Dabrowski, K. Saeed, W. Homenda, K. Saeed, W. Homenda, K. Saeed, W. Homenda, K. Saeed, A Multimodal Biometric User Identification System Based on Keystroke Dynamics and Mouse Movements. In Computer Information Systems and Industrial Management: 15th IFIP TC8 International Conference, CISIM 2016, Vilnius, Lithuania, September 14-16, 2016, Proceedings 15 (Springer International Publishing 2016), p. 672–681
17. C. Feher, Y. Elovici, R. Moskovitch, L. Rokach, A. Schclar, User identity verification via mouse dynamics. Inf. Sci. **201**, 19–36 (2012). https://doi.org/10.1016/j.ins.2012.02.066
18. M. Antal, N. Fej´er, Mouse dynamics-based user recognition using deep learning. Acta Universitatis Sapientiae. Informatica **12**(1), 39–50 (2020). https://doi.org/10.2478/ausi-2020-0003
19. N. Siddiqui, R. Dave, N. Seliya, Continuous authentication using mouse movements, machine learning, and Minecraft, (2021). arXiv preprint https://doi.org/10.48550/arXiv.2110.11080
20. B. Wang, S. Xiong, S. Yi, Q. Yi, F Yan, Measuring network user trust via mouse behavior characteristics under different emotions. In HCI for Cybersecurity, Privacy and Trust: First International Conference, HCI-CPT 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21 (Springer International Publishing 2019) pp. 471–481
21. C. Shen, Z. Cai, X. Guan, Y. Du, R.A. Maxion, User authentication through mouse dynamics. IEEE Trans. Inf. Forensics Secur. **8**(1), 16–30 (2012)
22. I. da Silva Beserra, L. Camara, M Da Costa-Abreu, M. Using keystroke and mouse dynamics for user identification in the online collaborative game League of Legends. 7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016), **8**, (2016). https://doi.org/10.1049/ic.2016.0076
23. M. Antal, L. Denes-Fazakas, User Verification Based on Mouse Dynamics: a Comparison of Public Data Sets. 2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI), 143–148 (2019). https://doi.org/10.1109/SACI46893.2019.9111596
24. T. Hu, W. Niu, X. Zhang, X. Liu, J. Lu, Y. Liu, An Insider Threat Detection Approach Based on Mouse Dynamics and Deep Learning. Security and Communication Networks **2019**, 1–12 (2019). https://doi.org/10.1155/2019/3898951
25. M. Antal, E. Egyed-Zsigmond, Intrusion detection using mouse dynamics. IET Biometrics **8**(5), 285–294 (2019)
26. I. Traore, I. Woungang, M.S. Obaidat, Y. Nakkabi, I. Lai, Combining Mouse and Keystroke Dynamics Biometrics for Risk-Based Authentication in Web Environments. Fourth International Conference on Digital Home **2012**, 138–145 (2012). https://doi.org/10.1109/ICDH.2012.59
27. S. Mondal, P. Bours, A study on continuous authentication using a combination of keystroke and mouse biometrics. Neurocomputing (Amsterdam) **230**, 1–22 (2017)
28. A. Subash, I. Song, I. Lee, K. Lee, AgeGen Bio Track: Continuous Mouse Behavioral Biometrics-Based Age and Gender Profiling in Online Education Platforms. In Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) **3**, 383–393 (2025). ISBN 978–989–758–737–5, ISSN 2184–433X

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.