Comparative Analysis of Machine Learning Models for Continuous Authentication

1. ABSTRACT

This paper explores the potential use of mouse dynamics as a continuous authentication method. Our study collects data from 15 participants playing the video game Roblox and analyzes 3 machine learning models where we perform a comparative analysis between them. Our results show promising accuracy rates of upwards of 97.51%, with small false negative rates as low as 1.32%, with k-Nearest Neighbors on average having the best performance. While further research is needed, our study suggests that mouse dynamics could be a useful tool in user authentication tasks. The novel contributions of our paper include providing a new mouse dynamics dataset and a comprehensive comparison of 3 machine learning models for continuous authentication.

2. INTRODUCTION

As online platforms and services grow, secure authentication is increasingly vital. Passwords and security questions can be vulnerable to cyber-attacks. Continuous authentication is another layer on top of passwords that verifies a user's identity throughout a session. It monitors behavior for signs of unusual activity or unauthorized access. Machine learning-based methods analyze user behavior patterns, identifying any anomalies indicating a security breach. These algorithms learn and adapt to new patterns in real-time, detecting threats before they cause damage. This prevents security incidents and data breaches. In this paper, we present a comparative study of machine learning-based continuous authentication methods. Our study is based on data collected from a popular online game, Roblox. We extracted features from the raw data and used them to train and evaluate our machine learning models. The models used in our study include K-Nearest Neighbors (KNN), Decision Trees (DT), and Support Vector Classifier (SVC). The novel contributions of this paper are as follows:

- Introduce a novel mouse dynamics dataset containing data from 15 users while they played the video game in Roblox. Dataset is available at: https://github.com/taxborn/mauth-research-project.
- Develop secure authentication models using k-Nearest Neighbors, Decision Tree, and Support Vector Classifiers, and comparing the results from these classifiers.

The remainder of this paper is organized as follows. Section 2 describes some related research for CA and AD. Section 3 provides a description of the data collection used in this research and the features extracted. Section 4 describes our approaches and classification techniques. Section 5 provides implementation and experimental results. Section 6 presents the discussion and analysis. Section 7 has limitations of this paper. Section 8 gives ideas for future works. Section 9 discusses the conclusions.

3. LITERATURE REVIEW

Our literature reference consisted of 18 papers. Of which, we as a team, read the first 2 papers and the other 16 papers were divided evenly among the team with 4 papers for each person. These papers later on helped us create literature review documents for reference when doing the project. These papers are in the area of continuous authentication schemes based on machine learning. Below are some of the papers out of the 18 that we as a team went through.

In one of the papers S. Fu, D. Qin [1] presented a combined neural network model (CNN-RNN) for mouse behavior-based user authentication. They used raw sequential mouse data as input. Moreover, they evaluated it on a real dataset which consisted of 15 users with 300 trials each. The result from their approach yields a 3.16% EER and a 99.39% AUC, with an authentication delay of 6.11 seconds on average. Wu, G. [11] talks about in his paper about getting 2.9% EER in their method. This shows the effectiveness of applying deep learning techniques for static mouse behavior-based user authentication.

In another paper L. Gao et al. [2] talked about two classifiers that are fused at the decision-level, which lessens their heavy reliance on training data. They used freely accessible Balabit dataset. Fifteen users' dynamic mouse data made up this dataset. Marcus Tan, Y. X. [12] wrote in their paper about the usage of the Balabit dataset as it is considered to have less occurrences of anomalies. The team breaks down the user's operation events into features like motion length, curvature, curvature, etc., and derives the user's behavior characteristics from the curve characteristics. 60 groups of mouse curves were used to extract 24.3% of the EER. It is possible to increase performance by 11.2% by using 3600 sets of mouse curves. Their approach combines the support vector machine optimized by genetic algorithm with the k nearest-neighbor algorithm to provide an error rate that is lower than that produced by the two approaches alone. Bours, P. [13] wrote about their results were action specific and are optimized using genetic algorithm.

Mondal, S. [3] talked about how they concentrated on developing a context-free continuous authentication system that responds to each individual action a user takes. They provide a reliable dynamic trust model algorithm that is adaptable to any continuous authentication system, regardless of the biometric modality. In another paper [13] Bours, P. described in the trust model that, they see each and every action leading to a change in trust and therefore, it is possible that each particular action lead to a lock out of the current user because one such action dropped the trust value below the lock out threshold. This dataset was gathered from 53 users using their data collection software in completely uncontrolled conditions. In order to avoid a scenario in which an attacker avoids detection by restricting to one input device because the system only checks the other input device, they considered both keystroke and mouse usage behavior patterns. The best finding of this study is that 50 out of 53 genuine users never have their accounts accidentally locked out by the system, while the remaining 3 genuine users (~ 5.7%) occasionally have their accounts locked out, typically after 2265 actions. In addition, only three out of 2756 impostors have evaded detection, or 0.1% of all impostors. After 252 actions on average, imposters are discovered.

4. DATASET

Our dataset consists of raw mouse dynamics data collected during the gameplay of a tower defense game on Roblox. The raw data collected includes the following features: ID, cursor X and Y positions, the button pressed, and button press duration.

4.1 DATA CLEANING

Data filtering techniques were employed to enhance the quality and reliability of the dataset. Duplicate entries and incomplete records that could potentially impact the processing and analysis were removed. This ensures that the dataset is cleansed and reliable.

4.2 DATA PROCESSING

In the data preprocessing step, we utilized a feature set that included various statistical, dynamic, and trajectory-based features. The extracted feature set used in our study consisted of the following features.

The mean, standard deviation, minimum, and maximum of X speed, Y speed, and overall speed; X acceleration. Y acceleration, and overall acceleration; speed or acceleration squared over distance. jerk, angle, tangent,

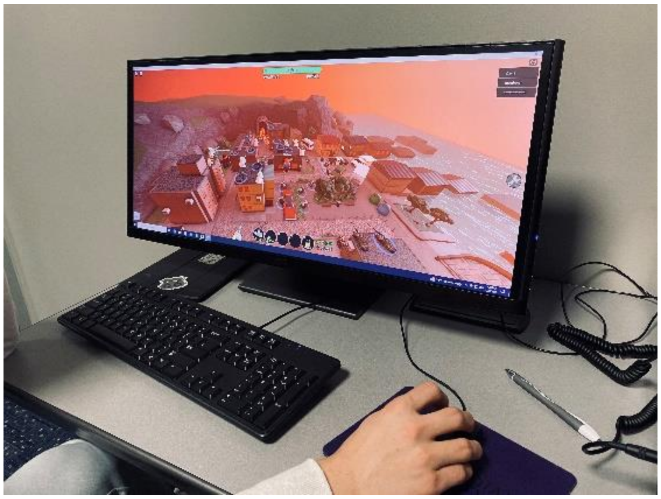

Figure 1. One of our subjects playing the game 'Tower Defense Simulator' from Roblox.

To preprocess the data, we first organized the raw data into sequences of 128 rows each, representing roughly 1 second of consecutive mouse dynamics data points on average per subject. Then, for each sequence, we calculated the above-mentioned features using appropriate formulas.

After calculating the features for each sequence, we further processed the data by normalizing or scaling the features, as needed, to ensure that they were on the same scale and to avoid any potential bias introduced by differences in feature magnitudes. This ensures that all features were treated equally during the subsequent analysis and modeling steps.

5. PROPOSED APPROACH AND METHODOLOGY

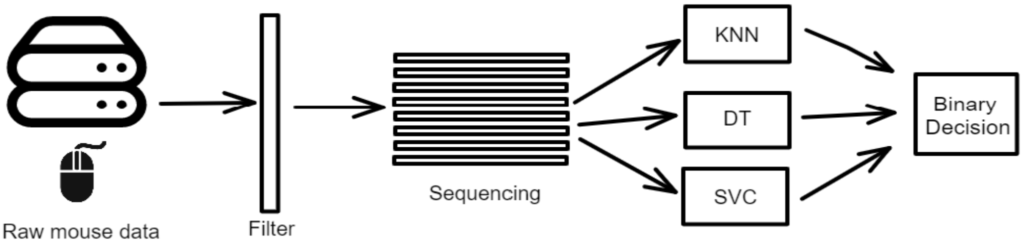

Figure 2. Model Diagram

This paper proposes a novel approach that involves testing the suitability of three machine learning algorithms: K-Nearest Neighbors (KNN), Decision Tree (DT), and Support Vector Classifier (SVC). Our primary objective is to evaluate each respective model and its potential viability in a continuous authentication model. Along with that perform a comparative analysis of our models against other papers.

5.1 K-NEAREST NEIGHBORS

The KNN algorithm was implemented using the scikit-learn library's K-Nearest Neighbors classifier, with hyper parameter tuning being done using GridSearch from the same library. It was found that using a k value of 3 and the Manhattan Distance metric were the best hyperparameters for this classifier.

5.2 DECISION TREE

The Decision Tree algorithm was implemented using the scikit-learn library's DecisionTreeClassifier class. Hyperparameter tuning was again performed using GridSearch where the maximum depth of 50 seemed to produce the best results. It was noted that Decision Tree has a fast prediction time at the expense of a slow fit time.

5.3 SUPPORT VECTOR MACHINE (SVC)

The Support Vector Machine (SVC) algorithm with a radial basis function (RBF) kernel was used as a classifier in this study to authenticate. The SVC algorithm was implemented using the scikit-learn library's SVC class, and hyperparameter tuning was performed using a grid search where the penalty parameter C was set to 100 and the kernel coefficient gamma was set to 'auto', where with the 55 features we used, evaluates to 0.018.

6. RESULTS

In this section we will discuss a comparative analysis of the 3 models we have trained. We have extracted a total of 55 features from the mouse for analysis, as mentioned in section . Our dataset contained a total of 1,246,577 events genuine user events, for an average of 83,105 events per subject.

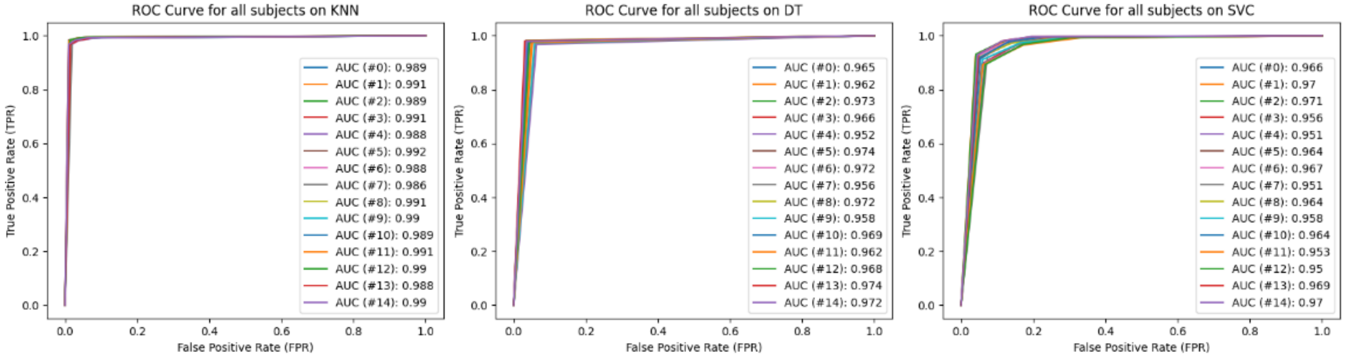

Figure 3. ROC curves for KNN, DT, and SVC models respectively

For evaluation, we split the training dataset into two different sets, one for training and one for testing. We trained the models on the training data, and tested using the testing set. We split the data using a standard 70-30 split, where 70% of the binary classified data was used for training, and 30% was used for testing. To measure the performance of each individual model, we used the following evaluation metrics: accuracy, F1-score, precision, recall, false positive rate (FPR), false negative rate (FNR).

Classifier	**KNN**	**DT**	**SVC**
Accuracy	97.49%	96.61%	95.51%
Recall	98.61	97.58	98.61
Precision	96.44	95.73	92.83
F1-Score	97.5	96.6	95.64

Table 1. Continuous authentication results for our models

Accuracy is the simplest metric to evaluate a model and is solely the percentage of correctly guessed sequences of events. According to Table 1, we were able to achieve accuracies ranging from 95.51% for SVC and up to 97.49% for KNN. False positive rate (FPR) is the rate in which an imposter user is authenticated as a genuine user, and false negative rate (FNR) is the classification of a genuine user as an imposter user. While aiming to reduce both FPR and FNR, we should aim to reduce FPR, since typically authenticating an imposter user is much more detrimental to a secure system than locking out a genuine user [7].

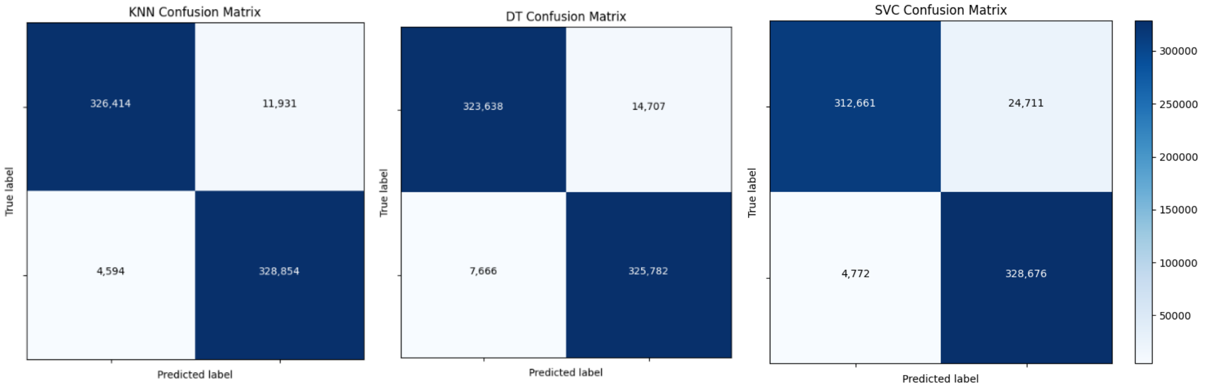

Figure 4. Confusion matrix metrics for KNN, DT, and SVC respectively.

We achieved a FPR of 3.68%, 4.51%, and 8.08% for KNN, DT, and SVC respectively. For FNR, we achieved 1.37%, 2.26%, and 1.32% for KNN, DT, and SVC respectively. In our models, KNN performed the best. For our KNN model, we were able to achieve an accuracy as high as 98% for user 5 in our model, and as low as 96.6% for user 7.

7. DISCUSSION AND ANALYSIS

The papers reviewed by the team suggest that continuous authentication systems based on mouse dynamics can achieve relatively low error rates with acceptable authentication delays. For example, the CNN-RNN model presented in [1] achieved a 3.16% EER and a 99.39% AUC with an authentication delay of 6.11 seconds on average. In Figure 3, we showcase an AUC ranging from 95% to 99%. Similarly, the continuous authentication system proposed in [2] achieved an error rate of 24.3%, which can be improved by using a larger dataset.

Our research shows that the mouse dynamics method for access control has a false rejection rate that nearly meets the European standard of 1%, indicating that it is effective at discriminating between users. However, the false acceptance rate is still above the required 0.001%, making the method unreliable as a standalone authentication method. It is important to note that authentication methods not meeting the European standard should not be solely relied upon, but combining mouse dynamics with other authentication methods could be valuable for user discrimination tasks.

Paper	Method	Results	Contribution
[1]	CNN-RNN Model with Mouse Movement	EER = 3.16%, AUC = 99.39%	Effective authentication time
[7]	SVM model with Mouse Movement	FRR = 0.86%	Fine-grained mouse movement metrics
[9]	NN with Keyboard Dynamics	EER = 2.13%	High accuracy in a far lower processing time
[10]	Naïve Bayes and SVM with Mouse and Keyboard	FPR = 2.10%, FRR = 2.24%	Multi-modal sensors with feature comparison

[11]	KNN, DT, and SVC with Mouse Movement	Accuracy = 54% – 95%	Lightweight mouse dynamics set to perform authentication
Our Research	KNN, DT, and SVC with Mouse Movement	Accuracy = 95.5% - 97.5%, FPR = 3.68% - 8.08%, FRR = 1.32% - 2.26%	ML model comparison with a novel mouse dynamics dataset

Table 2. Comparative analysis of our model evaluations versus other papers' model evaluations. When other papers have multiple results, ranges of their results are mentioned.

In conclusion, the literature review highlights continuous authentication using mouse dynamics as a promising research area that can lead to practical and effective authentication systems. Further research is required to develop more robust algorithms and feature sets that improve the accuracy and reliability of these systems.

8. LIMITATIONS

Some of the limitations that we faced are Our dataset only contains mouse movements from a mouse-intensive video game, and our results may not translate to day-to-day administrative tasks. For KNN, we noticed small False Acceptance Rates (FAR) and False Rejection Rates (FRR), of 0.0368 and 0.0137 respectively, however we would want to minimize FAR, as false positive authentications are much more detrimental to secure systems than false rejections. Our models were trained on data that all came from identical hardware, and our models may not perform the same if different subject data came from different computers.

9. FUTURE WORK

Future work could investigate the use of ensemble methods or other sets of dynamic features beyond mouse movement biometrics to further enhance system accuracy and robustness. Additionally, interpreting the decision-making process of the models through techniques such as feature importance analysis or model visualization could provide insights into the factors that contribute to the models' predictions and improve their decision making. It is also worth noting that in smaller tested feature sets the respective models' decision-making varied widely in the tens of percents. This suggests that there could be obvious variations in how some features map to a user differently based on their biometrics, demonstrating that specific features may be more helpful to specific subjects.

10. CONCLUSION

In conclusion, our research presents a comparative study of machine learning-based continuous authentication methods using a novel mouse dynamics dataset. Our study extracted features from the raw data and used it to train and evaluate machine learning models. Our results show that our k-Nearest Neighbor classifier is the most effective model for continuous authentication using mouse dynamics for our dataset and features selected.

In addition to the technical aspects of our research, it is important to consider the broader implications of our findings. Machine learning-based continuous authentication has the potential to significantly improve the security of online systems by providing near real-time authentication without disrupting the user's experience. Our research contributes to the development of these systems by providing a comparative analysis of three different machine learning models that can be used for continuous authentication based on mouse dynamics.

11. REFERENCES

[1] S. Fu, D. Qin, D. Qiao and G. T. Amariucai, "RUMBA-Mouse: Rapid User Mouse-Behavior Authentication Using a CNN-RNN Approach," 2020 IEEE Conference on Communications and Network Security (CNS), Avignon, France, 2020, pp. 1-9, doi: 10.1109/CNS48642.2020.9162287.

[2] c, "Continuous Authentication of Mouse Dynamics Based on Decision Level Fusion," 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 2020, pp. 210-214, doi: 10.1109/IWCMC48107.2020.9148499.

[3] Mondal, S., & Bours, P. (2017). A study on continuous authentication using a combination of keystroke and mouse biometrics. Neurocomputing, 230, 1–22. https://doi.org/10.1016/j.neucom.2016.11.031

[4] Antal, M., & Egyed-Zsigmond, E. (2019). Intrusion detection using Mouse Dynamics. IET Biometrics, 8(5), 285–294. https://doi.org/10.1049/iet-bmt.2018.5126

[5] Siddiqui, N., Dave, R. and Seliya, N. (2021) "Continuous User Authentication Using Mouse Dynamics, Machine Learning, and Minecraft," roc. of the International Conference on Electrical, Computer and Energy Technologies (ICECET) 9-10 December 2021, Cape Town-South Africa, doi: 10.1109/ICECET52533.2021.9698532.

[6] N. Zheng, A. Paloski, and H. M. Wang. (2011) "An efficient user verification system via mouse movements," in Proc. ACM Conf. Computer and Communications Security, Chicago, IL, 2011, pp. 139– 150.

[7] Marcus Tan, Y. X., Iacovazzi, A., Homoliak, I., Elovici, Y., & Binder, A. (2019). Adversarial attacks on remote user authentication using Behavioural Mouse Dynamics. 2019 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn.2019.8852414

[8] A.A. Ahmed, I. Traore, Biometric recognition based on free-text keystroke dynamics, IEEE Trans. Cybern. 44 (4) (2014) 458–472.

[9] L. Fridman, A. Stolerman, S. Acharya, P. Brennan, P. Juola, R. Greenstadt, M. Kam, Multi-modal decision fusion for continuous authentication, Comput. Electr. Eng. 41 (2015) 142–156.

[10] C.-C. Lin, C.-C. Chang, D. Liang, A new non-intrusive authentication approach for data protection based on mouse dynamics, in: Proceedings of the International Symposium on Biometrics and Security Technologies (ISBAST'12), IEEE, 2012, pp. 9–14.

[11] Wu, G., Wang, J., Zhang, Y., & Jiang, S. (2018). A continuous identity authentication scheme based on physiological and behavioral characteristics. Sensors, 18(2), 179. https://doi.org/10.3390/s18010179

[12] Marcus Tan, Y. X., Iacovazzi, A., Homoliak, I., Elovici, Y., & Binder, A. (2019). Adversarial attacks on remote user authentication using Behavioural Mouse Dynamics. 2019 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn.2019.8852414

[13] Bours, P., & Mondal, S. (2015). Performance evaluation of continuous authentication systems. IET Biometrics, 4(4), 220–226. https://doi.org/10.1049/iet-bmt.2014.0070