National College of Ireland

Continuous User Authentication on Mobile and Desktop Devices using Behavioural Biometrics – a Multimodal Approach

MSc Research Project
Cyber Security

Siddharth Sankar
Student ID: 23418541

School of Computing
National College of Ireland

Supervisor: Niall Heffernan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Siddharth Sankar

Student ID: 23418541

Programme: MSc Cyber Security **Year:** 2025

Module: MSc Research Project

Supervisor: Niall Heffernan

Submission Due Date: 28/01/2026

Project Title: Continuous User Authentication on Mobile and Desktop Devices using Behavioural Biometrics – a Multimodal Approach

Word Count: **7537** **Page Count 20**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Siddharth Sankar

Date: 27/01/2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Continuous User Authentication on Mobile and Desktop Devices using Behavioural Biometrics – a Multimodal Approach

Siddharth Sankar

23418541

Abstract

Traditional authentication methods are static and it's the year 2025. There is an exponential growth of mobile and desktop users, the vulnerabilities of these static methods have been exposed. The vulnerability this thesis aims to find a solution for is User's Session Hijacking. Continuous User Authentication (CUA) using behavioural biometrics can be a solution, but most approaches follow single modality that can struggle with high error rates due to noise and environmental factors, which has limited their widespread adoption. This thesis investigates the effectiveness of a Multimodal CUA framework that will fuse desktop keystroke data with touch gestures from a mobile to create a security layer across the user's devices.

This thesis adopted a quantitative experimental design, that pairs users from two different datasets to create 41 subjects with a unified identity. The analysis was done by doing a score-level fusion and using the Random Forest classifier to produce an Equal Error Rate.

The Results demonstrated a significant jump in security when fusing the two modalities together instead of using them independently. It was also found that the mobile only modality was insufficient to be used as a standalone authentication system since it achieved error rates of around 20% whereas the fused model achieved scores of around 5% showing a significant reduction of around 75% error rates. This thesis provides a blueprint for a scalable solution to Continuous Authentication using Behavioural Biometrics that enhances security while still maintaining user experience to promote widespread adoption.

Keywords: Continuous Authentication, Behavioural Biometrics, Fusion, Multimodal, Random Forest, Session Hijacking, Equal Error Rate, Security.

1. Introduction

For a long time, the foundational security approach in the digital world has relied upon static authentication methods such as passwords, face identification, fingerprint sensor and PIN codes. While they have been extremely beneficial and served their purpose of protecting our systems by being the initial gatekeepers. These methods however have significant risks associated with them. This is because they are limited to working at only the initial point of entry. This opens the system to unauthorised access that might occur during an active session, vulnerabilities such as phishing attacks, credential theft and other malicious activities can lead to serious consequences (Oduri, 2024).

To enhance security the field of Continuous User Authentication (CUA) is gaining popularity. The main aim of CUA is to passively and continuously verify a user's authenticity throughout their entire session. It does so by monitoring unique behavioural patterns such as typing

rhythms, touch gestures or mouse movements. It can also be done using various human traits such as a person's stride, it has been found that each person's stride is unique and contains distinctive characteristics (Mekruksavanich and Jitpattanakul, 2021). Such methods, however, are dependent on using additional IoT devices such as a smartwatch or any device with an accelerometer and gyroscope. This evolution from static to continuous verification is significant for cybersecurity. Nevertheless, the move of CUA from theoretical to practical applications have significant concerns regarding their accuracy, usability and adaptability to user variability such as fatigue and the general usability of such systems.

1.1. Motivation

CUA aims to solve the limitations in traditional authentication methods that create profound security risks. Passwords, Pins can be stolen through phishing, biometrics can be spoofed, systems left idle once logged in are also a significant vulnerability. Such risks point toward an urgent need for the implementation of effective security measures that persist beyond the initial login. As suggested by (Bours and Mondal, 2015) CUA systems that depend on a single behavioural biometric such as keystroke dynamics can achieve high accuracy in specific and controlled environments. The issue with such single modality systems is that they may fail if faced with factors such as user fatigue, change in environment or change in user's application use in the system. These factors can lead to high false rejection rates.
Furthermore, other authors have stressed the challenge of creating a unified identity profile in our multi-device world. The behavioural patterns on a desktop (keystroke and mouse) are fundamentally different from those on a mobile device (touch and swipe). This "cross-device gap" means that current frameworks often fail to create a single, unified identity profile that can reliably authenticate a user as they switch between platforms. These challenges highlight the need for a more holistic, adaptive, and multimodal approach that traditional, single-point security cannot provide
.

1.2. Research Objective

This study will describe an integrated, multimodal CUA framework that is tailored for cross-device environments. The framework aims to implement a strong level of security which uses disparate behavioural data – keystroke, mouse and touch. The objective is to build a framework in way which is continuous and does not sacrifice user convenience.

1.3. Research Aim

To create and evaluate an adaptive, preventative, and a strong CUA framework that can be used across desktop and mobile devices to increase the security of user's device while still maintaining seamless user experience. This thesis suggests the use of modern technologies such as advanced machine learning based anomaly detection models to detect threats in real and lock the imposter out. This will mitigate vulnerabilities such as spoofing, data breaches and unauthorized system access during any active session.
Machine learning will be used to develop a solution that identifies behavioral anomalies, that will help distinguish a genuine user from an imposter. The research also will aim to create an authentication system that will intelligently adjust the required level of authentication required in response to changes in user behavior such as user fatigue.

1.4. Research Questions

The founding question of the thesis is as follows, which can be split into three sub questions:
How can behavioural biometrics, when combined with traditional authentication methods, be adapted to enhance user security and promote widespread adoption across mobile and desktop devices?
RQ1: Can simple machine learning models effectively implement a fusion mechanism that improves the authentication accuracy when compared to a single-modality system?
RQ2: Will established security metrics such as EER, FAR and FRR improve in this multimodal approach for continuous authentication?
RQ3: What implications will the False Rejection Rate have on the user adoption?

2. Related Work

2.1. Introduction to Behavioural Biometrics

A major shift in user authentication is taking place, from the static one-time verification techniques to a more continuous monitoring. Upon conducting the academic literature review, it is seen that there is clear evolution that an identity of a user is defined in digital terms. Traditional approaches such as passwords rely on "what you know" or tokens – the "what you have ". On the other hand, the aim of behavioural biometrics is to try and focus on "who you are", this is done by looking into the subconscious patterns of how a user interacts with their device (Oduri, 2024).

These behavioural biometrics can be broadly categorised into desktop based – keystroke and mouse movements and mobile based – touch, gait and gyroscope modalities. A survey conducted by (Shadman *et al.*, 2023) points out that usage of single modalities have been able to achieve high accuracy but only in controlled environments. They suffer from a high false rejection rate (FRR) in the real world use due to factors such as change in environment and user fatigue. This thesis will critically analyse these modalities to provide a justification for the use of a multimodal approach using a score fusion that is being proposed in the thesis.

2.2. Desktop Modalities – Keystroke and Mouse Dynamics

2.2.1. Keystroke Dynamics

The keystroke modality is the most mature behavioural modality. It works on the principle of extracting various timing data between different key presses. The two major features are Dwell Time and Flight Time. Dwell time is the duration of a key press while flight time is the period between key release and next key pressed.

The foundational work on keystroke dynamics and its use for identification of user was done by (Monrose and Rubin, 2000), they proposed the use of a diagraph specific measure of variability and proposed the use of only arbitrary text instead of free text. This is a major limiting factor in today's need of continuous authentication. They also talked about how change in environmental parameters can affect the system's ability to authenticate the user. A more recent study by (Sae-Bae and Memon, 2022) argued that "intra – user variability" which is the change in a user's typing rhythm due to fatigue or external factors is only a solved problem theoretically. They demonstrate that even though high accuracy is possible which is a low False

acceptance rate, it comes at the cost of usability which implies a high False Rejection Rate. This insight is essential for this thesis since it supports the use of Random Forest classifiers, which

2.2.2 Mouse Dynamics

Mouse dynamics work by authenticating users based on their cursor velocity, angular velocity during point and click tasks and acceleration.

(Shen, Cai and Guan, 2012) notes that usage of mouse dynamics does not require any additional hardware like in the case of traditional biometrics such as fingerprint recognition or facial recognition. It suggests that mouse dynamics can be an important addition to the traditional use of behavioural biometrics, thus suggesting the use multiple modalities as the optimum solution of continuous authentication. It is also a good paper in terms of how one can mine the mouse movement data to understand and use patterns for evaluation of authenticity of user.

One critical limitation is highlighted by (Mondal and Bours, 2017), the issue with models being trained is that they are trained on one specific task for example a model might be trained on a user's mouse behaviour while browsing the net. This behaviour can change significantly when a user changes what they are doing for example playing a game or editing a word document. Such models can lead to a high False Rejection Rate thus locking a genuine user out mid-session. This weakness supports the thesis's objective of fusing the mouse data with keystrokes to mitigate such task dependant variations.

2.3. Mobile Modalities – Touch and Swipe Dynamics

In today's world it is a known fact that most users use both desktop and mobile devices. While keystrokes and mouse dynamics work on desktops, smartphones and tablets need something else. Current literature pivots us to Touch Dynamics.
(Frank *et al.*, 2013) proposed doing continuous authentication on mobile devices using 30 distinct features such as stroke length, pressure, area, etc. Their classifier was able to achieve an Equal Error Rate of around 4%. They however noted that their experimental findings showed that this cannot be used as a stand-alone authentication method but that it could be implemented along side static authentication to extend screen-lock time or as part of a multi-modal biometric authentication system.
(Papaioannou et al., 2023) points out that user acceptance of behavioural biometrics in mobile devices is affected by privacy fears. A user might feel data collection of behavioural features can lead to knowing the user's physical and mental state. Although this data is less sensitive than biometric data such as fingerprint, iris scan, etc. There is also lack of standardisation for behavioural biometric data such that there are no specific guidelines for storage of such data. This can lead to misuse of data by different providers that can use the data for their own purposes.

2.4. Fusion Strategies

As we see above all literature overwhelmingly points out that no single biometric is perfect, there is a need to use any behavioural biometric with conjunction of another behavioural biometric to properly secure the system and guarantee no harm to user experience. As many papers have suggested that the solution lies in a multimodal fusion. There is a debate on where this fusion must be done. It can be done at two levels:

i) Feature Level Fusion: At this level all raw data vectors are collected and fused simultaneously. (Ross and Jain, 2004) have argued that this method preserves the most information. One major draw back however in our case is that synchronous data is impractical to collect – typing on keyboard while using the mouse is impractical.

ii) Score Level Fusion: At this level all the output probabilities of independent models are combined. A very recent analysis by (Akintunde *et al.*, 2025) has provided compelling evidence that Score Level Fusion often matched with Feature Level Fusion performance in terms of accuracy achieved while reducing computational complexity. They made use of techniques such as Weighted Sum Rule and Product Rule. It was found that the simple weighted sum averages often outperformed complex fusion techniques in real world scenarios.

This finding helped us guide this thesis's adoption of the Simple Sum Rule Score Fusion technique, which will be kept simple and is able to balance performance with feasibility.

2.5. Summary and Gap Analysis

Deep Learning dominates most of the academic research being conducted in today's scenario. For example, (Tewari and Verma, 2022)'s Survey shows that many authors have used Deep Learning models such as CNN, CNN+RNN, Bi-LSTM, etc for Continuous authentication and have achieved high accuracy figures, but the issue of widespread adoption targeted by this thesis remains unsolved since they present significant barriers to computational power required. Below are some of the gaps found by conducting the literature review:

i) Computational Cost: As mentioned above most deep learning models such as the LSTMs and RNNs offer high accuracy but are expensive computationally speaking. They also require high resources for training and deploying. This makes them less suitable for widespread endpoint deployment compared to simple machine learning models such as KNN, SVM and Random Forest.
ii) No Cross-Device Unification: Almost all the research focuses on optimising continuous authentication on a single kind of device either desktop or mobile. There is lack of research that shows the advantage gained by fusing the two together for a user to make a single identity profile.
iii) Usability: The models come with major trade offs such as no free text being used for authentication. This is due to there being a high potential for increase in False Rejection Rates due to user fatigue or external factors.

This thesis fills the gap by providing a blueprint for a scalable and continuous authentication that addresses the above limitations moving away from single modality to a multi modal approach. The aim is to make use of this blueprint in future to create a rapidly deployable cross device security solution. As the literature suggests, we will be proving the efficacy of Score Level Fusion and finding out whether a low EER and an acceptable FRR for usability is achievable using existing non-synchronous datasets.

3. Research Methodology

3.1. Research Design and Framework

This thesis has adopted the quantitative experimental approach, where the objective is to do the evaluation of the efficacy of the multimodal fusion of behavioural biometrics. Since there are constraints around collecting synchronised cross-device data in real time, the research will utilize a simulation-based experimental approach that uses secondary data sources. The methodology will be beyond a simple observation of results; it will also include extensive testing of the hypothesis that Score level Fusion of different behavioural biometrics can meaningfully reduce the Equal Error Rate or EER compared to the single modality-based models as seen in the literature review.

3.2. System Environment

One of the main objectives of this thesis is to have widespread adoption of continuous authentication. This would be possible only if we created a lightweight solution by keeping the implementation constrained to standard consumer grade hardware.

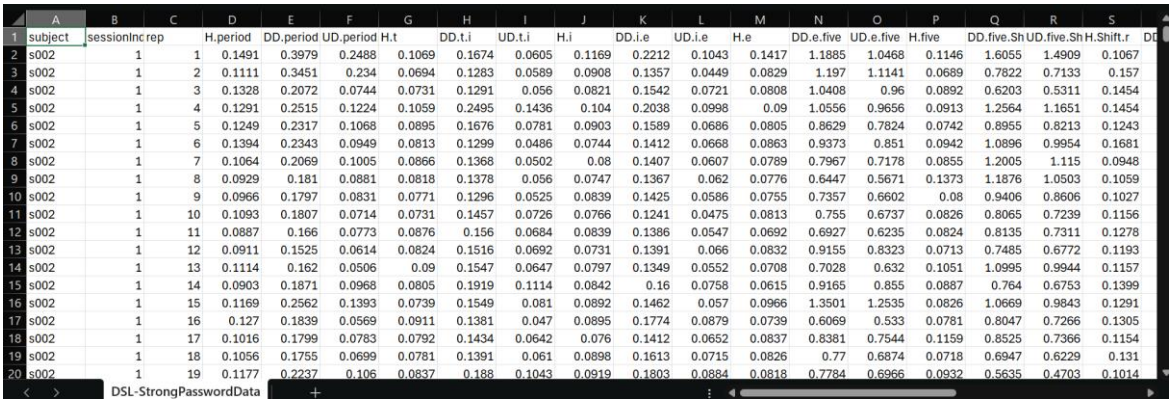
- Hardware Specifications: A laptop equipped with 16GB of RAM and a 11th Gen Intel(R) Core (TM) i7-11370H @ 3.30GHz processor was used to conduct experiments. No GPU acceleration was used which helped show that deployment was possible on standard endpoint devices.
- Software Used: Python 3.9 was used to implement the framework.
    - Pandas and NumPy was used to perform data manipulation.
    - Scikit-learn library was used to implement model due to its efficiency and industry standard algorithms.
    - Matplotlib and Seaborn were used for visualisation of results.
    - SciPy was used for calculating Equal Error Rate.

3.3. Data Description and Collection

As discussed in the literature review above, we require labelled benchmarked datasets of high quality to ensure that the evaluation has statistical high validity. In order to achieve this, we used the following datasets:

i) Desktop Modality: The CMU Keystroke Benchmark Dataset was selected as the standard to understand desktop behaviour.
    - Data Characteristics: (Killourhy and Maxion, 2009) had 51 subjects type a strong static password – (.tie5Roanl) on a laptop running windows xp and a software that recorded the keystrokes. The subjects typed the password 50 times each over 8 sessions. A gap of one day was given between each session. The password was typed 400 times by each subject by the end of collection.

- Data Structure: there are 31 columns that feature timing information such as hold time which implies time for which the key was held and time between when one key was released to the next key pressed.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	subject	sessionInc	rep	H.period	DD.period	UD.period	H.t	DD.t.i	UD.t.i	H.i	DD.i.e	UD.i.e	H.e	DD.e.five	UD.e.five	H.five	DD.five.Sh	UD.five.Sh	H.Shift.r
2	s002	1	1	0.1491	0.3979	0.2488	0.1069	0.1674	0.0605	0.1169	0.2212	0.1043	0.1417	1.1885	1.0468	0.1146	1.6055	1.4909	0.1067
3	s002	1	2	0.1111	0.3451	0.234	0.0694	0.1283	0.0589	0.0908	0.1357	0.0449	0.0829	1.197	1.1141	0.0689	0.7822	0.7133	0.157
4	s002	1	3	0.1328	0.2072	0.0744	0.0731	0.1291	0.056	0.0821	0.1542	0.0721	0.0808	1.0408	0.96	0.0892	0.6203	0.5311	0.1454
5	s002	1	4	0.1291	0.2515	0.1224	0.1059	0.2495	0.1436	0.104	0.2038	0.0998	0.09	1.0556	0.9656	0.0913	1.2564	1.1651	0.1454
6	s002	1	5	0.1249	0.2317	0.1068	0.0895	0.1676	0.0781	0.0903	0.1589	0.0686	0.0805	0.8629	0.7824	0.0742	0.8955	0.8213	0.1243
7	s002	1	6	0.1394	0.2343	0.0949	0.0813	0.1299	0.0486	0.0744	0.1412	0.0668	0.0863	0.9373	0.851	0.0942	1.0896	0.9954	0.1681
8	s002	1	7	0.1064	0.2069	0.1005	0.0866	0.1368	0.0502	0.08	0.1407	0.0607	0.0789	0.7967	0.7178	0.0855	1.2005	1.115	0.0948
9	s002	1	8	0.0929	0.181	0.0881	0.0818	0.1378	0.056	0.0747	0.1367	0.062	0.0776	0.6447	0.5671	0.1373	1.1876	1.0503	0.1059
10	s002	1	9	0.0966	0.1797	0.0831	0.0771	0.1296	0.0525	0.0839	0.1425	0.0586	0.0755	0.7357	0.6602	0.08	0.9406	0.8606	0.1027
11	s002	1	10	0.1093	0.1807	0.0714	0.0731	0.1457	0.0726	0.0766	0.1241	0.0475	0.0813	0.755	0.6737	0.0826	0.8065	0.7239	0.1156
12	s002	1	11	0.0887	0.166	0.0773	0.0876	0.156	0.0684	0.0839	0.1386	0.0547	0.0692	0.6927	0.6235	0.0824	0.8135	0.7311	0.1278
13	s002	1	12	0.0911	0.1525	0.0614	0.0824	0.1516	0.0692	0.0731	0.1391	0.066	0.0832	0.9155	0.8323	0.0713	0.7485	0.6772	0.1193
14	s002	1	13	0.1114	0.162	0.0506	0.09	0.1547	0.0647	0.0797	0.1349	0.0552	0.0708	0.7028	0.632	0.1051	1.0995	0.9944	0.1157
15	s002	1	14	0.0903	0.1871	0.0968	0.0805	0.1919	0.1114	0.0842	0.16	0.0758	0.0615	0.9165	0.855	0.0887	0.764	0.6753	0.1399
16	s002	1	15	0.1169	0.2562	0.1393	0.0739	0.1549	0.081	0.0892	0.1462	0.057	0.0966	1.3501	1.2535	0.0826	1.0669	0.9843	0.1291
17	s002	1	16	0.127	0.1839	0.0569	0.0911	0.1381	0.047	0.0895	0.1774	0.0879	0.0739	0.6069	0.533	0.0781	0.8047	0.7266	0.1305
18	s002	1	17	0.1016	0.1799	0.0783	0.0792	0.1434	0.0642	0.076	0.1412	0.0652	0.0837	0.8381	0.7544	0.1159	0.8525	0.7366	0.1154
19	s002	1	18	0.1056	0.1755	0.0699	0.0781	0.1391	0.061	0.0898	0.1613	0.0715	0.0826	0.77	0.6874	0.0718	0.6947	0.6229	0.131
20	s002	1	19	0.1177	0.2237	0.106	0.0837	0.188	0.1043	0.0919	0.1803	0.0884	0.0818	0.7784	0.6966	0.0932	0.5635	0.4703	0.1014

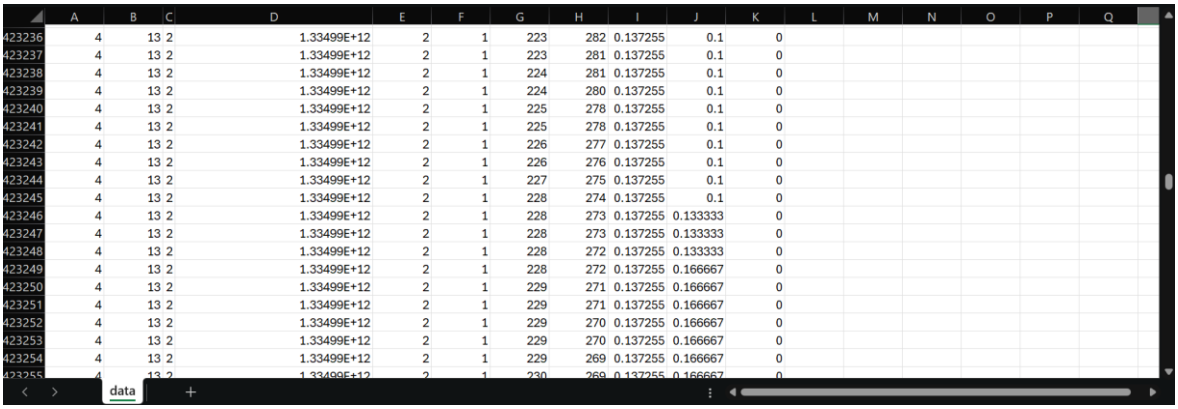
DSL-StrongPasswordData

Figure 1 CMU Keystroke Data

- Another advantage of using this data set is that it captures inter user variability or fatigue since the subjects were made to type the password over multiple sessions.

ii) Mobile Modality: the TouchAnalytics Dataset was used for the mobile behavioural biometrics.

- Data Characteristics: (Frank *et al.*, 2013) had 41 subjects perform tasks such as scrolling and navigating on android phones such as Samsung Galaxy, Nexus 1 and Nexus S.
- Data Structure: For each touch stroke 30 features were extracted. Features include touch up, down, finger movement on screen. Also, the x-coordinate, y—coordinate, area covered are some of the other data points that were collected.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
423236	4	13	2	1.33499E+12	2	1	223	282	0.137255	0.1	0						
423237	4	13	2	1.33499E+12	2	1	223	281	0.137255	0.1	0						
423238	4	13	2	1.33499E+12	2	1	224	281	0.137255	0.1	0						
423239	4	13	2	1.33499E+12	2	1	224	280	0.137255	0.1	0						
423240	4	13	2	1.33499E+12	2	1	225	278	0.137255	0.1	0						
423241	4	13	2	1.33499E+12	2	1	225	278	0.137255	0.1	0						
423242	4	13	2	1.33499E+12	2	1	226	277	0.137255	0.1	0						
423243	4	13	2	1.33499E+12	2	1	226	276	0.137255	0.1	0						
423244	4	13	2	1.33499E+12	2	1	227	275	0.137255	0.1	0						
423245	4	13	2	1.33499E+12	2	1	228	274	0.137255	0.1	0						
423246	4	13	2	1.33499E+12	2	1	228	273	0.137255	0.133333	0						
423247	4	13	2	1.33499E+12	2	1	228	273	0.137255	0.133333	0						
423248	4	13	2	1.33499E+12	2	1	228	272	0.137255	0.133333	0						
423249	4	13	2	1.33499E+12	2	1	228	272	0.137255	0.166667	0						
423250	4	13	2	1.33499E+12	2	1	229	271	0.137255	0.166667	0						
423251	4	13	2	1.33499E+12	2	1	229	271	0.137255	0.166667	0						
423252	4	13	2	1.33499E+12	2	1	229	270	0.137255	0.166667	0						
423253	4	13	2	1.33499E+12	2	1	229	270	0.137255	0.166667	0						
423254	4	13	2	1.33499E+12	2	1	229	269	0.137255	0.166667	0						
423255	4	13	2	1.33499E+12	2	1	230	269	0.137255	0.166667	0						

data

Figure 2 TouchAnalytics Mobile Data

- This dataset provides a significant advantage to our thesis since unlike the static keyboard dataset, this dataset represents a continuous behaviour. This noisy data was required to test the toughness of the Random Forest Classifier used.

3.4. Data Preprocessing and Feature Extraction

The raw data collected is bound to be noisy since its behavioural based. In order to get the best results, the following preprocessing of data was done.

3.4.1. Desktop Data - Feature Extraction and Normalisation

As seen above the data contained raw time stamps, from which two key features were extracted:

- Dwell Time: this is the duration of which a key is pressed i.e. $(T_{release} - T_{press})$.
- Flight Time: this is the gap between two consecutive key presses i.e. $(T_{press}^{n+1} - T_{release}^{n})$.
- Normalisation: typing speeds is something that varies from user to user. To mitigate this Standard scaling (Z-score Normalisation) was applied to the dataset. This helped ensure the model learnt from the average speed of typing of a user rather than the absolute speeds.

3.4.2. Mobile Data Preprocessing and Scaling

The TouchaAnalytics data had to be pre-processed for missing values, and some touch features required scaling. Below are the steps performed on the raw data:

- Mean Imputation: To prevent data loss missing values were handled using mean imputation, that is they were replaced with statistical mean of that feature for the particular user.
- Scaling: Touch features have different units such as pixels vs milliseconds. Therefore, a Min-Max scaling was applied to range from 0-1, since values with large magnitudes such as coordinates would have skewed the decision trees.

3.5. Machine Learning Implementation

3.5.1. Algorithm Selection

Random Forest was selected for this framework, based on a preliminary comparative study done on the CMU Keystroke benchmark dataset. In the experiment three machine learning algorithms were compared: Support Vector Machines (SVM), k-Nearest Neighbours and Logistic Regression. The following observations were made:

- Logistic Regression: this was the worst performing algorithm with an accuracy of only 80%, it failed to capture the non-linear relationship between the features. It had an EER of 6.8%.
- SVM: this was a significant improvement over logistic regression but still proved to be highly sensitive to outliers. This is an issue for behavioural biometrics since it is common to have outliers in such data. SVM had an accuracy of 86.3% and an EER of 0.4%.
- KNN: Although it had a lower accuracy of 85.5% it still had a better EER of 0.3%. the issue with KNN was the computational cost and latency. This will make it unsuitable for real time data.

Learning from these results and to address the limitations seen, the random forest ensemble method was chosen for this thesis. It will address the limitations in the following ways:

- Variance Reduction: by using the ensemble method, aggregation of multiple decision trees will minimise the sensitivity to outliers as seen in SVM.

- Non-Linear: it will be able to capture features in the dataset that were missed by logistic regression due to its linear nature.
- Speed: Random Forest unlike KNN will perform better for real time data, thus satisfying the user usability issue.

The Random Forest classifier is robust to noise and has great execution speed compared to deep learning models (Apruzzese *et al.*, 2018), he also notes that Random Forest is efficient for large datasets like ours and excel at multiclass problems. One thing to note is that it may lead to overfitting due to deeper trees.

3.5.2. Training Strategy

A one vs rest approach was used to train the model. Binary classification was done for each target user. Class 1 was assigned to Genuine User and Class 0, or impostor was assigned to a random sample of data from all other users.
The data was split 80/20 for training/testing. This ensured that the ratio of genuine to impostor sample remained consistent across the model.

3.6. Fusion

Since we used to different independent datasets for desktop and mobile that contained different individuals, we had to simulate a single user using both devices by doing the following:

- Mapping: a mapping function was randomised to pair user A from CMU Keystroke dataset to a user B from the TouchAnalytics dataset. This helped simulate a single user C using both devices.
- Score Level Fusion: as seen in the literature review above unlike feature level fusion which would need synchronous timestamps, a score level fusion on the other hand combines the probability of outputs of the independent models. (Ulery *et al.*, 2006) have described a linear method - the weighted sum rule which we apply in the following way:
    - $\boldsymbol{S_{final} = (w \times P_{desktop}) + ((1 - w) \times P_{mobile})}$
      Where $\boldsymbol{P_{desktop}}$ is the probability from the keystroke data and $\boldsymbol{P_{mobile}}$ is the probability from the touch analytics dataset. **w** or the weight is 0.5 to signify equal contribution.

3.7. Evaluation Techniques

Keeping in mind the Security vs Usability argument we saw above, the following evaluation techniques have been used to see the performance:

i) False Acceptance Rate (FAR): this will give the measure of security our model is capable of. It is calculated using

$$FAR = \frac{False\ Positives}{False\ Positives + True\ Negatives}$$

ii) False Rejection Rate (FRR): it is the measure of how usable the model is for users. It is calculated using

$$FRR = \frac{False\ Negative}{False\ Negatives + True\ Positives}$$

iii) Equal Error Rate (EER): this is the point where the FAR is equal to the FRR. These metrics are considered to be the best way to evaluate biometric data. A low EER would mean the model is working well to classify the genuine users from the imposters.

iv) ROC Curve: this will help visualise the separation abilities of the classifier.

4. Design Specification

The framework was developed keeping in mind the scalability and maintainability of the model. The proposed solution is designed to act as a Multimodal Continuous Authentication Framework. The architecture has four main components in the pipeline – Data Ingestion, Data Processing, Classification and Fusion. These components are modular in nature; this helps with maintainability and allows individual modules to be updated without harming the framework. For example, the touch module can be replaced with a different modality if required. The framework works in the following way:

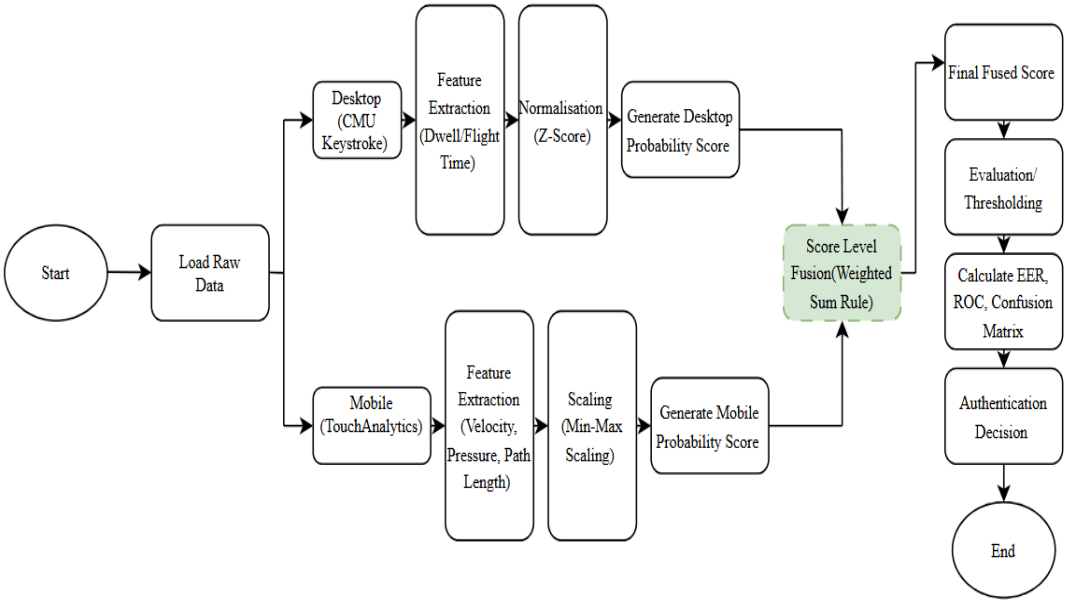


Figure 3 Implemented Framework

i) Ingestion Layer: this layer is responsible for loading the raw keystroke data from desktop and touch features from mobile devices. The input here is raw CSV files that contain keystroke timestamps from desktop, and the touch data contains touch coordinates, pressure, orientation of finger, etc.

ii) Transforming Layer: this is where the preprocessing will take place. The desktop data will be normalised using Z-Score standardisation and the touch data will be scaled using min-max scaling to keep the pressure and velocity readings in range. scaling. For this thesis since the datasets we use are from two different sets of users, we will also map users to simulate single user in this layer. This will help create virtual subjects that stimulate one user using both mobile and desktop.

iii) Classification Layer: we will make use of Random Forest ensemble classifier here. The output will be independent confidence scores for each modality. In order to let the authentication function even if one system fails, we will model the two modalities independent of each other.
iv) Fusion Layer: Using weighted sum rule here, the final layer will perform a score level fusion. Here the probability confidence score will get aggregated and make a universal trust score. The weight is kept at w = 0.5 so that both devices can be treated at par. This can be changed for mobile or desktop depending on noise.

The design is made in a way that will ensure enhancement in user security while also promoting widespread adoption. Below are some of the functional and non-functional requirements that will make sure this is achievable:

i) Latency: the system must be capable of processing the input vectors under a second to maintain usability and a seamless user interaction.
ii) Security: A target of less than 5% EER is set for the final fused output.
iii) Hardware Requirement: the architecture is intended to run on a standard consumer level CPU and does not require any specialised GPU acceleration.

5. Implementation

The above flowchart gives an overview of the steps followed and is followed with a detailed explanation of each step.

5.1. Implementation Tools and Environment

Python 3.9 programming language was used to implement the final solution, Google collab environment was used to leverage cloud-based computational capabilities. The core development relied on multiple open-source libraries:

- NumPy and Pandas: these libraries were used to manipulate and preprocess the raw data. These tools were primarily responsible for the pairing logic of users from the two datasets.
- Scikit-learn: this was used for the machine learning part of the solution. It was essential for three main tasks, implementation of Random Forest Classifier, train – test data splitting and calculating the probability estimates for the final evaluation of the model.
- Matplotlib and Seaborn: these libraries were used in the generation of visualisations of the final output. Mainly the ROC curve and Confusion Matrix.
- SciPy: this was used to calculate the FRR, FAR and EER.

5.2. Feature Engineering

1) To get usable features from the raw mobile data a custom engine for feature extraction was developed. The following features were calculated using the below methods:
   i) Average Velocity: to understand muscle memory, this will correspond to the speed of the user's swipe

$$v_{average} = \frac{\Sigma\, distance}{Time_{end} - Time_{Start}}$$

ii) Average Pressure: to get a metric of the user's physical interaction with their mobile.

iii) Path Distance: This feature was used to distinguish between an efficient swiper from a user with an erratic swipe pattern. The end-to-end distance was compared to the total sum of path length to understand the curve of the swipe.

2) The mobile data we used was in pixels for distance while the keystroke data was in milliseconds for time. Since we are going to be doing fusion, it was important to normalise and scale this data.

i) The Z score method was applied to normalise the Keystroke data.

ii) A min-max scaling was applied to the Touch Mobile data, this helped bound the velocity and pressure within (0,1), this helped stability in Random forest's tree splitting process.

5.3. Produced Outputs

There were mainly three categories of outputs generated in the implementation phase:

i) Data Artefacts: the raw data was transformed into structured matrices that were suitable for further processing. Here are the three Data Artefacts outputs:

- Normalised Desktop Data: to overcome variations in typing speed, the CMU benchmark dataset was transformed using the Z-Score technique.
- Mobile Dataset Features Extraction: An extraction script was used to process the raw data that had 912,133 rows. The output of the script resulted in a structured matrix that contained 20,682 valid strokes from 41 users. The raw coordinates data for each user was then transformed into behavioural features – Duration of stroke, Average pressure, average velocity and distance.
- Paired Dataset: this was the consolidated output that contained 14.997 samples where each sample was a representation of virtual user that had both keystroke and touch attributes thus resulting in data that could be fed into the classifier.

ii) The intelligence of the system comprised of two separate machine learning models:

- Rf_Desk: the ensemble random forest classifier using n = 100 estimators was trained on the samples. the model's output was configured to produce a score $\boldsymbol{P_{desktop}}$(Probabilistic Confidence Score) rather than a binary class label.
- Rf_Mob: this ran parallel to Rf_Desk and trained on the extracted touch features above. Similar to Rf_Desk, this also outputs a score $\boldsymbol{P_{mobile}}$ which represents the likelihood of the user's biometric behaviour being genuine.

iii) Fusion Output: implementation of the Score – Level Fusion strategy happens in this phase. This module takes the independent scores of Rf_Desk and Rf_Mob and applies the Weighted Sum Rule to them:

$$P_{Fused} = (0.5 \times P_{desk}) + (0.5 \times P_{Mob})$$

The final output P_Fused is a single confidence vector for the testing test and determines the authenticity of the user.

iv) The final stage of the implementation produces a quantitative report to validate our hypothesis, it includes ROC curve visualisations, confusion matrix and the exact EER numbers of each independent model and the fused model. The research questions it answers are:

- Was there any security enhancement by using a multimodal approach.
- Was usability affected.
- Did it have latency for real time use.

6. Evaluation

6.1. Findings Overview

The research's primary objective was to test the ability of a multimodal authentication system to be able to secure cross-device user sessions using a score level fusion framework. The results from the model's evaluation revealed a hierarchy of performance. Below are the Results achieved and their analysis.

6.2. Results And Analysis

The proposed framework and the authentication model was evaluated under three scenarios to understand how fusion benefits the multi modal continuous authentication. Below we will see how the model performs in terms of quantitative metrics such as the EER and ROC curves we achieved using the Random Forest Classifier.

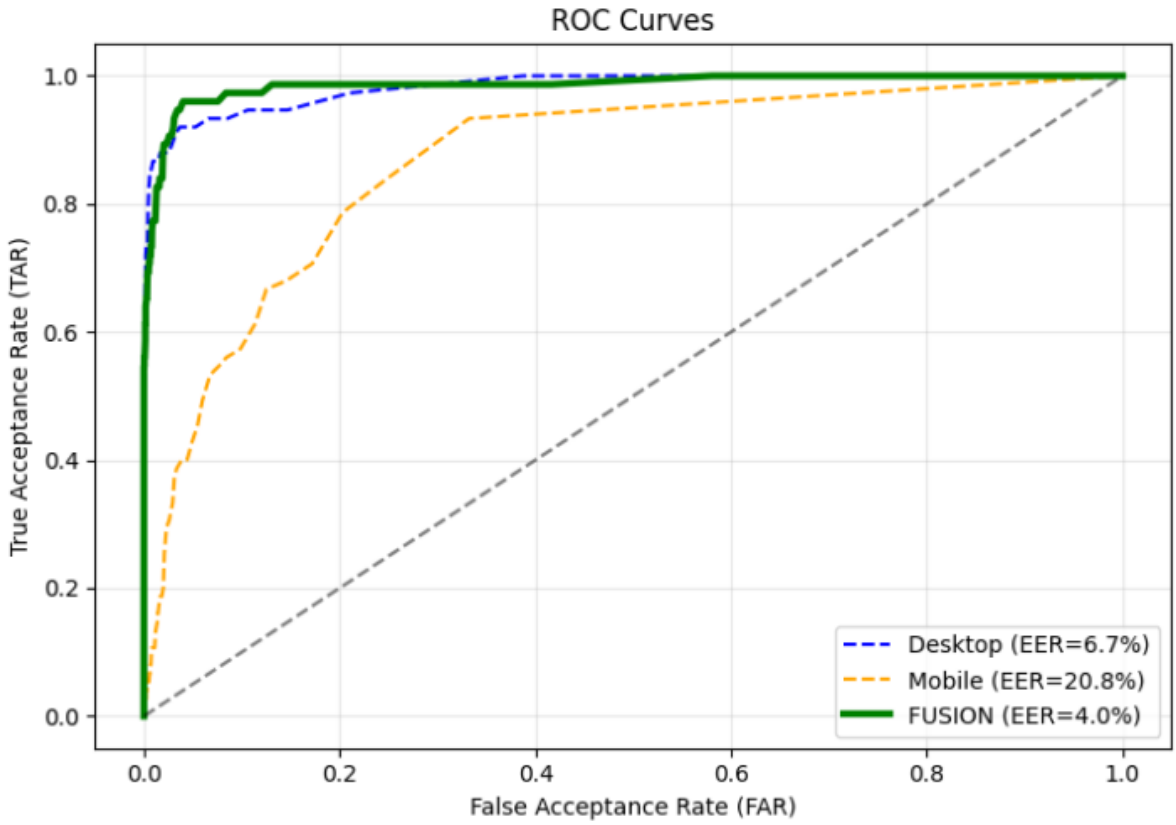


Figure 4 ROC Curves Obtained

6.2.1. Observations from the ROC Curve

- The orange line which is of low quality is of the Mobile Only model, the EER obtained is 20.8%. This signals that the TouchData is quite noisy thus preventing the model from learning the user's behaviour perfectly.
- Desktop is the blue line which is smoother and stronger and is a good result for desktop modality.
- The Green line is for the fused model and even after the noise from the mobile modality it has extracted useful signals to lower the overall EER of the authentication system.

6.2.2. Case 1: Mobile Only Authentication

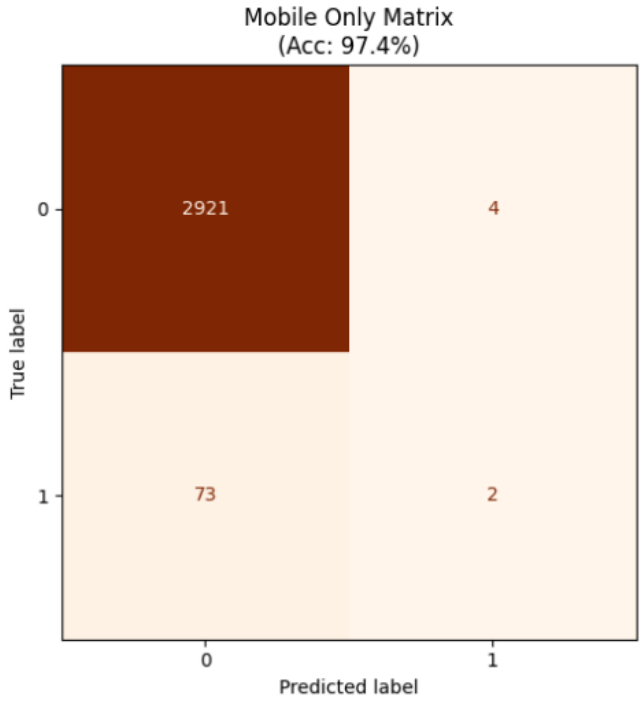


Figure 5 Confusion Matrix for Mobile Only Modality

- In this scenario, only the TouchAnalytics Dataset is used, after extracting Duration of stroke, Average pressure, average velocity and distance from the raw data. The result is as shown below
- EER of 20.78% was achieved.
- **Analysis:** the trajectory of the ROC curve indicates that the Random Forest struggled with classifying a genuine user from an impostor. The high error rate of approximately 1 in every 5 attempts is justified since we are dealing with mobile behavioural biometrics which is naturally volatile. Multiple factors can cause this volatile behaviour such as, variation in finger posture, screen smear, variation in pressure, etc. due to this a considerable amount of noise is added into the data.
- this adds on to the hypothesis that for continuous authentication stand alone modalities are insufficient to improve security, a 20.78% EER is significantly higher than industry standards.
- The confusion matrix shows that the genuine user got accepted only 2 times and got rejected 73 times, meaning the model is under-confident. It allowed the impostor only 4 times while locking them out 2921 times.

6.2.3. Case 2: Desktop Only Authentication

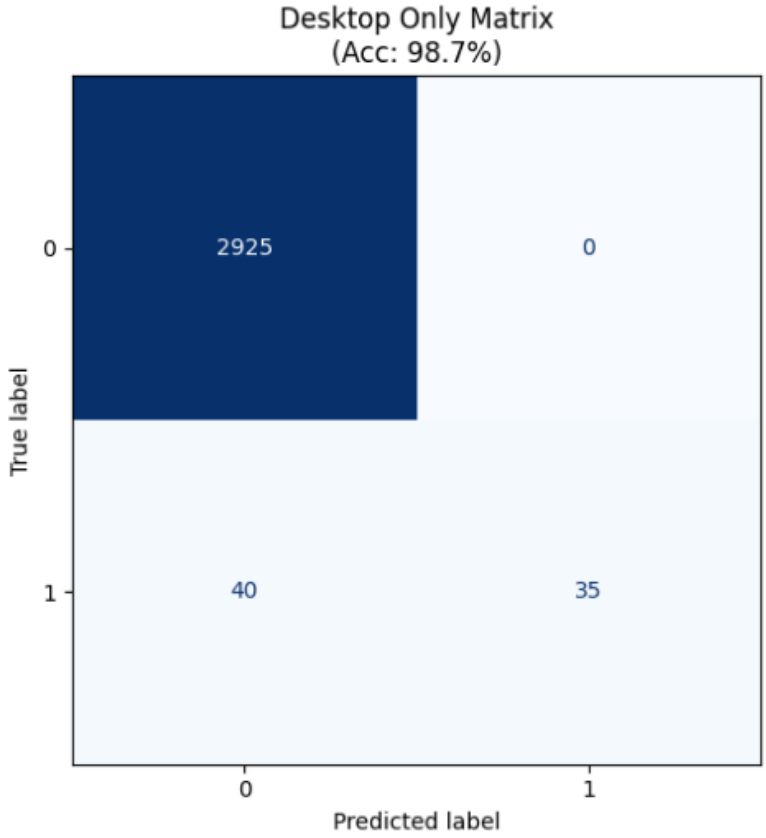


Figure 6 Confusion Matrix for Desktop Only Modality

- In this scenario, only the CMU Keystroke Dataset was used in which the two main features used were dwell time and flight time. The result achieved is shown below:
- EER of 6.67% was achieved.
- **Analysis:** here the random forest classifier was able to create a tighter decision boundary. This can be attributed to the nature data. Typing is more rhythmic which reduces the noise in the data.
- The result is great since it validated that the continuous authentication in the desktop environment is performing well. But as discussed above this cannot be relied upon since once the user steps away from the desktop to use their mobile phone, this necessitated the fusion approach.
- The confusion matrix shows that the impostor user got accepted 0 times which is a great result. However, the genuine user got rejected 40 time which is cumbersome for the user.

6.2.4. Case 3: Multimodal Fusion Authentication

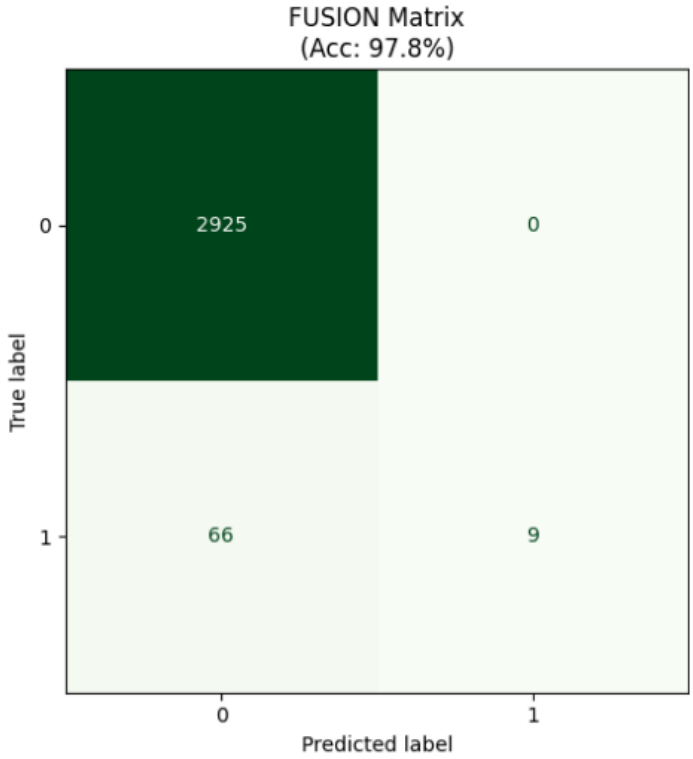


Figure 7 Confusion Matrix for Fused Modality

- In this scenario we have implemented, the score level fusion architecture that combines the probabilities from Case 1 and Case 2. The result achieved is shown below:
- EER of 4% was achieved.
- **Analysis:** There was a significant improvement in the security as seen by the EER percentage. An absolute improvement of 16.78% was seen over the Mobile Only model. It is also lower than the desktop only model signifying that combining both produced better results. This is an indication that the fusion is having a synergetic effect. It also means that even though the mobile data was noisy, it contributed enough to the keystroke data to reduce the over all EER.
- Case three proves the original hypothesis of the thesis that a multimodal fusion approach successfully increases the security. It also showed that the high trust desktop environment extended to the low security mobile environment.
- The confusion matrix here shows that it maintains perfect security and that no impostors get through although the genuine user is still rejected 66 times. This is due to the threshold of 0.5 being too strict. In real-world deployment this threshold needs to be lowered to reduce the false rejection rate.

The below table provides a Summary of the Results:

Case	Modality	EER	FAR	What it Means
Case 1	Touch	20.78%	Very High	The system is vulnerable to impostors.
Case 2	Keystroke	6.67%	Low	Strong Performance
Case 3	Fused	4%	Lowest	Cross Device coverage and enhanced security.

Table 1 Result Summary

6.3. Analysis With Reference to Research Questions

The improvement seen in case 3 is not just a decrease in percentage but an optimal number, operations wise. To understand this better, let's extrapolate the results to understand them better, for example in a real-world deployment if there are 1000 log in attempts, Case 1 or mobile only would accept or reject approximately 209 users – this is a lot speaking both usability and security wise. The third case reduces this error number to just 40 users. This correlates to enhancing user experience which will promote widespread adoption, as a genuine user will be able to use this authentication system without getting locked out often while still maintaining security.

RQ1 was to check if a simple machine learning method such as the Random forest Classifier would be able to effectively implement a Fusion mechanism for continuous authentication using behavioural biometrics. The evidence seen above answers the question, when the weighted sum rule using random forest is applied the resulting EER score is more robust than the individual systems as seen by the ROC Curve in Figure 4. Statistically it is of significance since the fusion mechanism simply did not produce the average EER of the two systems combined but instead improved the overall score. This demonstrated that simple algorithm that is computationally efficient like the Random Forest is capable of learning decision boundaries for heterogeneous ingested data and can classify between genuine users and impostors with an acceptable error rate.

RQ2's aim was to be able to quantify the security enhancement of the multi-modal approach, the fused EER result of 4% represented a substantial jump in the security compared to 20.78% in the mobile only model. The risk reduced therefore from a 1 in 5 chance to a 1 in 25 chance of an impostor taking control of the session. For an impostor to take control of the session they must not only replicate the keystroke pattern but also the touch behavioural pattern to stay logged into the user's system.

Adoption of security technologies is often hindered due to poor user experience such as such as a genuine user getting incorrectly blocked, RQ3 focused on the adoption of continuous authentication using behavioural biometrics. As we see a user would be blocked 1 in 5 times if they only used touch data for continuous authentication. Fusion enables usability since it drastically reduces the number of times a genuine user would get locked out while still maintaining security.

6.4. Comparison with Previous Research

The results produced for the thesis align with and go on to extend the existing knowledge of behavioural biometrics. In this section we will validate how significant our findings are when compared against the literature reviewed in Section 2.

i) Limitations of Mobile Modality: The finding that mobile modality results in a high EER of 20.78% Resonates with the limitations that were identified by (Frank *et al.*, 2013). Although they achieved an EER of less than 4%, they did so by making use of highly engineered features in a controlled setting. They also came to the same conclusion as this thesis that mobile modality cannot be used as an authentication method on its own but can be used in a multimodal setting. This resonated the need of using a fusion approach.

ii) Confirming Efficacy of Score-Level Fusion: Our fusion framework was a success, resulting in a 4% EER. This provided the experimental support to the recent analysis done by (Akintunde *et al.*, 2025). They argued that Score-Level performance can match Feature-Level performance while reducing the computation need. Our implementation proves this hypothesis. We used a simple weighted sum rule with random forest probabilities and achieved a strong security framework. We did this without requiring complex resources which (Ross and Jain, 2004) noted can be a significant barrier for a feature-level fusion.

iii) Addressed the Deep Learning Obstacle: (Tewari, 2022) highlights that the field now is being dominated by deep learning models which although offer high accuracy but come with significant barriers to widespread adoption with the major reason being the computational costs. This thesis addressed this gap by showcasing how simple machine learning random forest classifier can do the job as seen by the competitive 4% EER number we achieve.

6.5. Discussion

6.5.1. Overview

The thesis had a clear objective of evaluating if a Multimodal Continuous User Authentication is feasible security widespread adoption wise. The results we see above, provide substantial evidence of how our proposed model performs keeping in mind the objective. Upon interpreting the figures obtained we gain several key insights.

i) There is clearly a high error rate, when processing the TouchAnalytics Data by itself, which confirms the need for a multi-modal system to overcome the instability of touch behavioural biometrics data.

ii) The keystroke data when processed by itself provides a strong foundation as one of the sources of data to prove a genuine user.

iii) The fusion model's superior performance confirms that even in the presence of one weaker system in a multi modal authentication model, the combined results prove effective when it comes to securing the system as seen by the EER numbers, meaning that it can leverage the strength of one strong biometric channel to compensate the weakness of another.

iv) The use of Random Forest classifier also proves effective when it comes to the processing speed, thus becoming one of the factors that help in adoption of Continuous Authentication.

6.5.2. Experiment Critique and Suggestions on Improvement

i) A major limitation of the current design is its reliance on training on a static set of data. User behaviour can adapt and change over time. It might differ due to factors such as fatigue, change in mood or something as small as change in phone case can affect the model. This may cause an increase in the False Rejection Rate over time. A deployable solution should include a constant learning mechanism. This will help the model train periodically with new samples and get to understand user behaviour better with existing knowledge. It will help reduce the FRR over time.

ii) This thesis had to make use of pairing users from two different datasets. This is crucial since we had to use different behaviours from two different datasets and combine them. A change in user's behaviour will be seen across both modalities in real world scenarios. To improve upon this synchronised data collection must be prioritised.

7. Conclusion and Future Work

The primary aim of the thesis was to overcome the vulnerability of user session hijacking by answering the question "How can behavioural biometrics, when combined with traditional authentication methods, be adapted to enhance user security and promote widespread adoption across mobile and desktop devices?"
This was achieved by developing a Multimodal Continuous User Authentication framework. The thesis's objectives were to transform disparate behavioural data types by implementing Score-Level Fusion, measuring the security enhancement and how this method of CUA affects widespread adoption. We combined the CMU Keystroke dataset and the TouchAnalytics Dataset. To implement this, we used the Random Forest Ensemble Classifier on the paired dataset which contained 41 subjects. The thesis was successful in meeting all the defined objectives. The eventual Equal Error Rates we achieved by testing all three models – Mobile Only, Desktop Only and Fusion were 20.78, 6.67%, and 4% respectively showing a gradual increase in security. The FAR and FRR numbers achieved showed that such a framework can be adopted with ease.

The future work needs to focus on real-world deployment which has challenges such as:

i) Dynamic Fusion: in the current model we used a fix weight of 0.5 while using the simple weighted sum principle for fusion. In the future this weight needs to adjust dynamically based on the user's activity such as is the user walking or is stationary, this can be done using sensors on the phone such as a gyroscope and an accelerometer. This will help the system reduce the dependence on the mobile data during those noisy activities while still maintaining security using keystroke data.

ii) For this thesis, we implemented the model on Google collab, further research should be done by deploying the model locally and seeing how the performance is and whether this is commercially viable.

iii) Privacy of user's data must also be prioritised; encryption must be added to further secure the data or process the data on user's device to protect the user's data.

iv) There is also the need to research the advancement of how artificial intelligence can be used in simulating user's behaviour and if yes, how can this model be optimised to protect from such advancements in technology.

References

Oduri, S. (2024) "Continuous Authentication and Behavioral Biometrics: Enhancing Cybersecurity in the Digital Era."

Bours, P. and Mondal, S. (2015) "Continuous Authentication with Keystroke Dynamics," in Y. Zhong and Y. Deng (eds.) *Gate to Computer Science and Research*. 1st ed. Science Gate Publishing P.C., pp. 41–58. Available at: https://doi.org/10.15579/gcsr.vol2.ch3.

Mekruksavanich, S. and Jitpattanakul, A. (2021) "Biometric User Identification Based on Human Activity Recognition Using Wearable Sensors: An Experiment Using Deep Learning Models," *Electronics*, 10(3), p. 308. Available at: https://doi.org/10.3390/electronics10030308.

Monrose, F. and Rubin, A.D. (2000) "Keystroke dynamics as a biometric for authentication," *Future Generation Computer Systems*, 16(4), pp. 351–359. Available at: https://doi.org/10.1016/S0167-739X(99)00059-X.

Sae-Bae, N. and Memon, N. (2022) "Distinguishability of keystroke dynamic template," *PLOS ONE*, 17(1), p. e0261291. Available at: https://doi.org/10.1371/journal.pone.0261291.

Shen, C., Cai, Z. and Guan, X. (2012) "Continuous authentication for mouse dynamics: A pattern-growth approach," in *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2012)*. *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2012)*, pp. 1–12. Available at: https://doi.org/10.1109/DSN.2012.6263955.

Mondal, S. and Bours, P. (2017) "A study on continuous authentication using a combination of keystroke and mouse biometrics," *Neurocomput.*, 230(C), pp. 1–22. Available at: https://doi.org/10.1016/j.neucom.2016.11.031.

Frank, M. *et al.* (2013) "Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication," *IEEE Transactions on Information Forensics and Security*, 8(1), pp. 136–148. Available at: https://doi.org/10.1109/TIFS.2012.2225048.

Ross, A. and Jain, A. (2004) "Multimodal biometrics: An overview," pp. 1221–1224.

Akintunde, O.A. *et al.* (2025) "Comparative analysis of score level fusion techniques in multi-biometric system," *LAUTECH Journal of Engineering and Technology*, 19(1), pp. 128–141.

Apruzzese, G. *et al.* (2018) "On the effectiveness of machine and deep learning for cyber security," in *2018 10th International Conference on Cyber Conflict (CyCon)*. *2018 10th International Conference on Cyber Conflict (CyCon)*, pp. 371–390. Available at: https://doi.org/10.23919/CYCON.2018.8405026.

Ulery, B. *et al.* (2006) *Studies of biometric fusion*. 0 ed. NIST IR 7346. Gaithersburg, MD: National Institute of Standards and Technology, p. NIST IR 7346. Available at: https://doi.org/10.6028/NIST.IR.7346.