eXplainable AI(XAI) for Touch-Stroke Biometrics: Insights from SHAP

Soodamani Ramalingam
School of Physics, Engineering and Computer Science
University of Hertfordshire
Hatfield, Hertfordshire, UK
https://orcid.org/0000-0001-5005-5809

Dominic Lovric
Algorise Ltd
London, UK
dominic@algorise.co.uk

Ooi Shih Yin
Faculty of Information Technology (FIST), Multimedia University, Jalan Ayer Keroh Lam Melaka, 75450
syooi@mmu.edu.my

Richard Guest
Professor of Biometric Engineering
University of Southampton, UK
r.m.guest@soton.ac.uk

Moises Diaz
Physics Department
Universidad de Las Palmas de Gran Canaria, Spain
moises.diaz@ulpgc.es

Fabio Garzia
Safety & Security Engineering Group - DICMA, SAPIENZA
University of Rome, Italy
fabio.garzia@uniroma1.it

David Lawunmi
IT Services
Queen Mary University of London, UK
d.lawunmi@qmul.ac.uk

***Abstract*— This paper presents an XAI-based framework for touch-stroke behavioral biometrics. Initially, a Random Forest classifier is trained to perform user classification, and feature importances are derived from the model's internal metrics. Subsequently, SHAP explanations are applied to obtain model-agnostic feature attributions. A comparison between the two approaches is then conducted to identify consistent patterns of feature relevance, informing the decision to exclude redundant or less influential features. The findings underscore the potential of integrating XAI into behavioral biometrics to enhance transparency and user trust.**

***Keywords*— eXplainable AI (XAI), touch-stroke dynamics, biometrics, SHAP**

I. Introduction

Biometric authentication systems are increasingly valued for providing secure and user-friendly identification methods. Among various biometric modalities, touch-stroke dynamics are emerging as a promising approach, capturing unique behavioural patterns from user interactions with touch devices. This modality is particularly well-suited for continuous authentication in high-security contexts such as mobile banking and healthcare.

In parallel, the field of eXplainable AI (XAI) is gaining momentum, aiming to make AI model decisions transparent and interpretable. Within biometric authentication, XAI plays a critical role in addressing the black box nature of machine learning models, fostering trust and accountability among users and stakeholders. This is especially relevant as regulatory and ethical frameworks increasingly demand explainability in AI-driven systems.

A generic biometric authentication system incorporating Explainable AI (XAI) is shown in Fig. 1. Such a system integrates two interrelated concepts: *explainability*, which refers to understanding an algorithm's internal design, training, and decision-making process; and *interpretability*, which focuses on presenting those insights in a human-understandable, cause-and-effect manner.

One emerging behavioural biometric modality that stands to benefit from XAI is touch-stroke biometrics. This

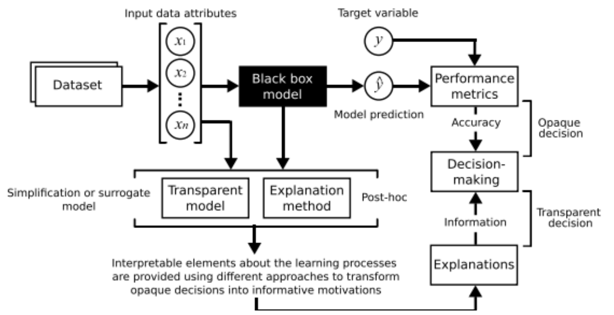


Fig.1 Explainability and Interpretability for a Machine Learning Model[1]

technique analyses how users' swipe, tap, or interact with touchscreens to authenticate their identity. It is a relatively recent area of research, gaining traction due to the increasing ubiquity of touchscreen devices. Touch strokes are defined as sequences captured by a device's touch sensors during user interaction [2]. Due to the distinctiveness of human musculoskeletal structures, individuals produce unique movement patterns [3] allowing for the extraction of digital signatures from touch points or keystrokes captured through built-in sensors [4]. Prior work in this area includes mobile signature verification using handheld devices [5], often examining the influence of data acquisition methods and classifiers such as Hidden Markov Models (HMMs) across different datasets. Frank et al. [4] explored K-Nearest Neighbours (KNN) and kernel-based Support Vector Machines (SVM), while Serwadda et al. [6] carried out a benchmark of 10 different classifiers. These studies collectively highlight the potential of touch gestures as a reliable biometric modality.

To the best of our knowledge, there has been no prior work in eXplainable AI (XAI) specifically addressing touch-stroke dynamics. By focusing on the nuances of touch dynamics and the unique behavioural traits they capture, we can enhance the transparency, trust, and usability of biometric authentication systems. We introduce methods for feature attribution, visualisation, and user-specific explanations that build upon existing XAI techniques, tailored specifically for the touch-stroke domain. This contribution is a first step towards providing insights for robust solutions in the field of touch-stroke biometric authentication.

The rest of the paper is organised as follows: In Section II, we introduce a model-agnostic explanation technique namely SHAP. In addition, touch-stroke biometric model is described for SHAP to generate explanations of the classifier. Section III provides a detailed explanation of the results generated against the ground truth and validated by cross correlation and manual feature importance extractions. Section IV shows classifier performance improvement through a pruned set of features. Section V provides conclusion and further work.

II. METHODOLOGY

A. SHAP – Local Interpretability of Predicitve Models

SHapley Additive exPlanations (SHAP) [7] provides a well-established approach for assigning a numerical value to represent each feature's contribution to a specific model prediction. The model output is interpreted as a deviation from a baseline value, typically the average prediction across the dataset. The sum of all feature contributions, expressed as Shapley values, reconstructs the model's output for a given instance. The Shapley value for a feature i is given by:

$$\varphi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|\,!\,.\,(|F| - |S| - 1)!}{|F|!} \left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S\left(x_S\right) \right]$$

where,

- F is the full set of features,
- S is any subset of F that does not include feature i,
- $f_{S\cup\{i\}.}\left(x_{S\cup\{i\}}\right)$ is is the output when feature i added to S,
- φ_i quantifies the average contribution of feature i across all possible subsets.

The difference $f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S\left(x_S\right)$ represents the added influence of feature i when combined with the feature set S. The weighting factor $\frac{|S|!\,.(|F|-|S|-1)!}{|F|!}$ reflects the number of permutations where subset S appears before feature i in the ordering of features. Calculating the exact Shapley values involves evaluating the model on a large set of feature combinations which can become computationally expensive as the number of features increases [15]. To address this, approximation algorithms are commonly used. Kernel SHAP [6] is a widely used model-agnostic method that estimates Shapley values through weighted linear regression. For models with a tree-based structure, such as the Random Forest classifier used in our work, a more efficient algorithm called Tree SHAP [7] is suitable and provides exact values for this class of models.

In this study, we apply SHAP to explain the decisions of a machine learning classifier used for touch stroke biometric authentication. We cross-verify feature contributions from Random Forest's traditional feature importances and gain a more reliable understanding of how model outputs are formed.

B. Touch-Stroke Biometrics

In this section, we consider the biometrics pipe-line for touch-stroke dynamics.

1) Antal Dataset. This study utilises the Antal dataset[8], a touch screen interaction database comprising 231,371 user-device interaction entries. These were collected from 71 users over a four-week period using eight different mobile devices, including both tablets and smartphones with varying screen sizes. Two tasks were involved:

(i) **Reading activity**, involving vertical scrolling strokes through text.
(ii) **Image gallery navigation**, requiring horizontal swipes to select a favourite picture.

The dataset is organised into three main files:

- **raw_data.csv**: Contains stroke-level touch information, including device_id, user_id, timestamp, action (0 = touch down, 1 = touch up, 2 = touch move), touch coordinates, pressure, and finger area.
- **devices.csv**: Lists technical specifications for each device, including device_id, screen density, resolution, OS version, and screen DPI values (x-dpi, y-dpi).
- **users.csv**: Provides user demographic data and information on touchscreen experience levels.

2) Pre-processing and Feature Engineering

The Antal dataset displays a pronounced orientation bias of mobile devices held by users: 226,058 entries (97.7%) were recorded in portrait mode, compared to just 5,313 entries (2.3%) in landscape mode. This imbalance reflects natural user preferences and ergonomics, as participants were allowed to use their devices freely, mirroring everyday mobile usage. However, the disparity presents challenges when comparing touch dynamics across orientations. To address this, the dataset was split by device orientation, allowing tailored analyses and recognising that user interaction behaviour can differ significantly between portrait and landscape modes.

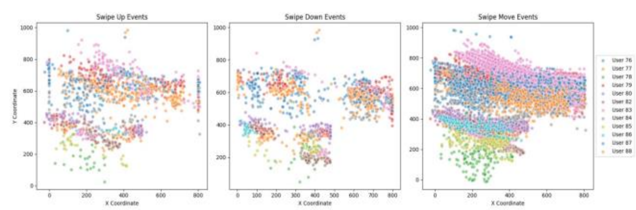
Following practices in recent studies on touch-stroke dynamics [9, 10] a series of pre-processing steps were implemented to ensure data quality and consistency, including missing values detection, outliers removal removing 15% of the original data, data normalisation and balancing datasets using under sampling to avoid model bias[11]. An initial sampling adequacy analysis revealed an uneven distribution of swipes across devices. Using pivot table analyses (samples/device and users/device), and further refining by samples/activity and users/activity, a working subset was established consisting of 65 swipes per user across 10 devices.

For feature engineering, we adopted two key methods to enrich the data with dynamic interaction metrics. Drawing on insights from [12] which highlights influential features in touch-stroke dynamics, we derived both velocity and acceleration features at the level of individual touch points and entire swipes. These included:

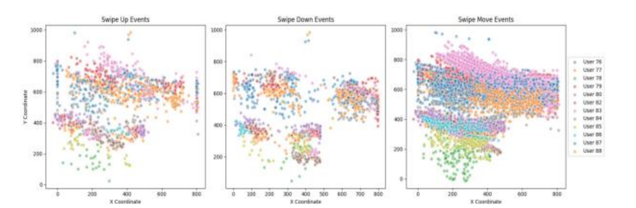
- Point-to-point velocity and acceleration,

- Overall swipe velocity and acceleration. Median velocity of the last three points of each swipe (velocity_last_3_pts),
- Median acceleration of the first five points of each swipe,
- 20th, 50th, and 80th percentiles of pairwise velocity and acceleration,
- Deviations from the end-to-end swipe line, measured at the same percentiles (20th, 50th, 80th).

3) Touch-Stroke Dynamics of Users. In the context of this study, variations in user swipe behaviour across tasks and devices based on sensor-derived and engineered features are used to interpret user interactions. Fig.2 illustrates swipe patterns of 12 users during two activities when device is held in portrait mode. In the **image gallery task** (Fig. 2a), swipe-up and swipe-move events are widely distributed, reflecting diverse user interactions, while swipe-down events cluster lower on the screen, likely due to thumb reach. In the **reading task** (Fig. 2b), swipe behaviour is more structured: swipe-up endpoints are concentrated, swipe-downs are centrally initiated, and swipe-move trajectories are predominantly linear and vertical. These task-specific patterns highlight the contextual nature of touch dynamics, with implications for adaptive interfaces and behavioural biometrics.



a) Image Gallery Viewing

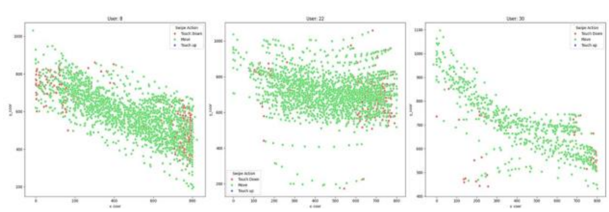


b) Reading Activity

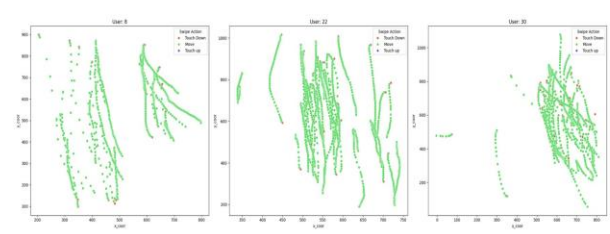
Fig. 1. Swipe Pattern of Selected Users Across All Devices Based on Activity (Portrait Mode)

Similarly, Fig.3 compares swipe trajectories across activities on the same device. During **gallery browsing** (Fig. 3a), swipes exhibit a strong horizontal spread, reflecting lateral navigation, with users varying between clustered and dispersed patterns. In **reading** (Fig. 3b), swipes are predominantly vertical, aligning with scrolling, with swipe length and occasional horizontal shifts indicating reading pace and re-engagement. These contrasts underscore the behavioural richness and task-adaptive nature of touch interactions, supporting user modelling and authentication.

4) Feature Correlation. We establish a baseline using a Random Forest classifier to assess feature importance across portrait and landscape modes. Orientation-specific differences emerge, with some features consistently ranked high, though discrepancies limit definitive conclusions. Complementary correlation analysis (Fig. 4) reveals strong associations especially among swipe distance, velocity, and vertical coordinates, highlighting the influence of interaction geometry. A recurring negative correlation between swipe duration and velocity across both modes underscores the role of temporal dynamics. These insights form a foundation for subsequent explainability assessments, linking model behaviour to orientation- and task-dependent user traits.

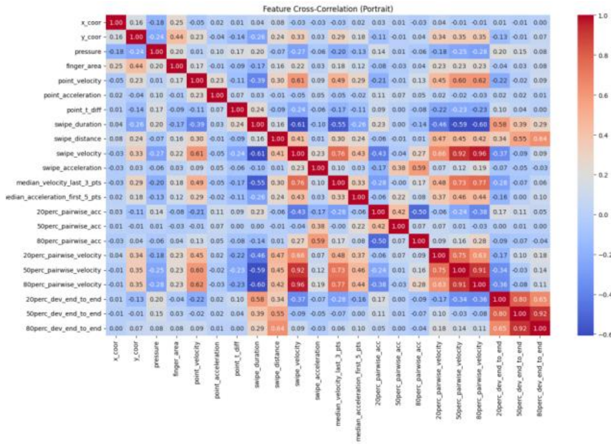


a) Image Gallery Viewing

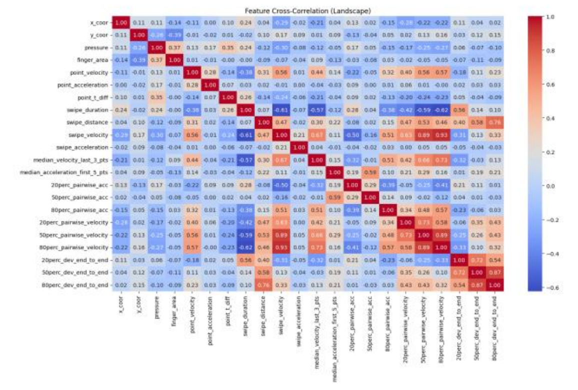


b) Reading Activity

Fig. 2. Swipe Patterns of Users on the Same Device



c) Portrait Mode



d) Landscape Mode

Fig.4. Cross Correlation Feature Matrices

5) Initial Model Performance. Random Forest models were trained on balanced subsets of the data for both portrait and landscape orientations. The initial results showed promising accuracy, with the portrait model achieving an accuracy of approximately 91.21% and the landscape model achieving approximately 82.93%.

III. XAI Model Interpretations

In this Section, we present the application of SHAP to interpret the Random Forest models trained on the touch-stroke dynamics data. We analyse explanation outputs for both correctly and incorrectly classified samples from the combined portrait and landscape orientations to uncover feature influences on model decisions.

A. *SHAP Interpretations*

The primary use of SHAP is to explain the feature importance and contributions of different features to the predictions made by the trained Random Forest models for both the balanced portrait and balanced landscape datasets. The Mean Absolute SHAP Importance (MASI) for a feature is the average magnitude of the SHAP values for that feature across all the instances in the sample test set and averaged across all possible output classes as shown in Figs. 7(a) and 7(b). Features with higher MASI are considered more important by the SHAP method because they have a larger average impact on the model's output. Key inferences made are:

- **Key Discriminators:** `finger_area`, `pressure` and `swipe_distance` have high SHAP importances in both plots, it reinforces the idea that features such as these are consistently important characteristics for user identification, regardless of orientation.
- **Orientation-based Importance:** Features that are ranked high in the Portrait plot but lower in the Landscape plot are more important for user identification in portrait mode and vice-versa. Relative importance of features such as `x_cord` and various velocity/acceleration metrics, shifted significantly between orientations.

B. *Local Interpretability with SHAP*

As a next step, we analyse specific instances of correctly and misclassified test samples using SHAP and compare the RF and SHAP importances numerically. Using these findings will guide feature removal. The natural next step would be to evaluate the performance of the models after removing the low-importance features to see if this simplification improves or maintains accuracy. Table I shows the top 5 features with the largest positive and negative impact on the prediction for each specific instance. In effect, it reveals the most influential features in the model's decision for a swipe.

For a **correctly classified instance**, the features with large positive SHAP values are the ones that strongly supported the correct user's class. Features with negative SHAP values for the correct class might have suggested a different user, but their combined impact was outweighed by the positive contributions.

For a **misclassified instance**, the features with large positive SHAP values are those that strongly supported the *incorrect* predicted user's class. Features with negative

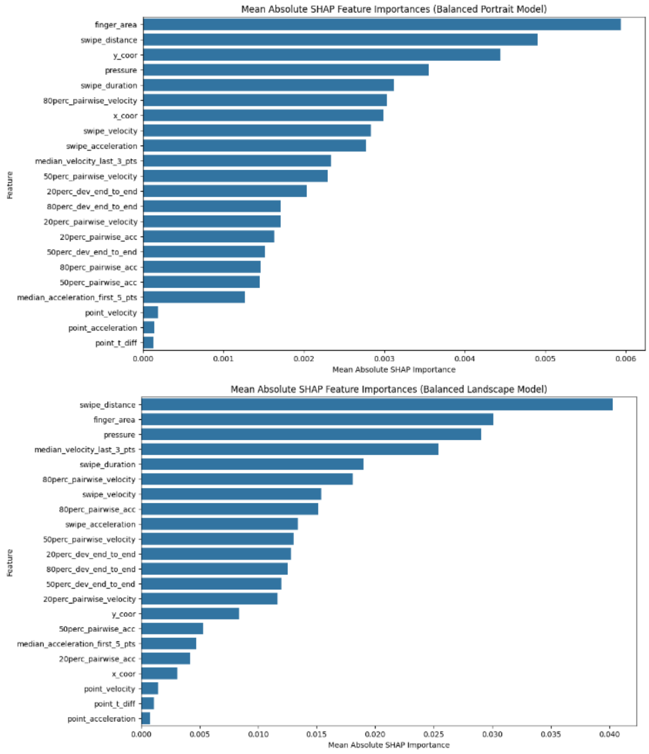


Fig. 7. SHAP Summary plots of overall feature importances for the models separated by orientation. Top-Portrait, Bottom-Landscape

TABLE I. Test Instances for SHAP Local Interpretation

Actual Vs Predicted Class	Orientation	Positive Impact Features	Importance	Negative Impact Features	Importance
(63,63)	Portrait	median_vel x_cord 80pw_vel swipe_acc swipe_vel	0.05 0.05 0.05 0.04 0.04	pt_acc pt_t_diff pt_vel median_acc_first_5pts pressure	-0.0007 0.00009 0.0036 0.0126 0.0152
(75,71)	Portrait	80pw_vel swipe_dist swipe_vel swipe_acc swipe_dur	0.03 0.03 0.03 0.02 0.02	20pw_vel x-cord 20pw_acc pt_acc 80pw_acc	-0.006 -0.003 -0.002 -0.002 -0.002
(1,1)	Landscape	swipe_dist finger_area pressure swipe_vel 20pw_acc	0.17 0.16 0.13 0.11 0.07	x-cord pt_t_diff pt_acc pt_vel 20pw_acc	0.000 0.000 0.000 0.000 0.002
(63,35)	Landscape	swipe_acc swipe_dur swipe_vel 50pw_vel swipe_dist	0.25 0.10 0.07 0.03 0.03	50pw_acc y_cord x_cord med_vel_last_3pts 50pc_dev_end2end	-0.003 -0.001 -0.00007 -0.0006 0.0011

SHAP values for the predicted class were pushing the prediction away from that incorrect user, but their impact was not strong enough to overcome the positive influences towards the wrong class.

Comparing the features and their SHAP values between correctly and misclassified instances can help understand why the model made an error in the latter case. For example, if a feature that usually has a strong positive SHAP value for a user's correct class has a negative or much smaller positive value in a misclassified instance of that same user, it might indicate an unusual swipe pattern for that user or a pattern that is very similar to another user for that specific feature.

Mean Absolute SHAP Values (MASV): The Mean Absolute SHAP value for a feature represents its average magnitude of impact on the model's output across the sampled dataset. A higher value means the feature has a greater overall influence on the model's predictions. Table II shows the Mean Absolute SHAP Importances for both Portrait and Landscape orientations (sorted by importance within each orientation). We can infer that the MASV importances give us a good overall sense of which features the models are leveraging most heavily to distinguish between users in each orientation, averaged across the dataset.

- **Consistency:** `finger_area` and `swipe_distance` appear to be important in *both* orientations, suggesting they are generally strong discriminators regardless of how the phone is held. pressure also seems relatively important in both.
- **Orientation Differences:** As expected, `y_cord` is more important in Portrait than in Landscape, while `x_cord` appears somewhat more important in Landscape (though not as dominant as `y_cord` in Portrait). Several velocity and acceleration feature also show shifts in importance between orientations.
- **Lower Importance Features:** Features like `point_velocity`, `point_acceleration`, and `point_t_diff` consistently have very low MASV in both orientations and therefore consider them for removal.

TABLE II. MASV IMPORTANCES FOR THE DATASET

Portrait Orientation		Landscape Orientation	
Top Features	**MASV X 10^{-2}**	**Top Features**	**MASV X 10^{-2}**
finger_area	-0.59	swipe_dist	-0.03
swipe_dist	-0.49	80pw_vel	-0.03
y_cord	-0.44	finger_area	-0.03
pressure	-0.36	swipe_dur	-0.02
x_cord	-0.30	pt_t_diff	-0.001

C. Comparison of MASV with traditional RF feature importances

Let's compare the normalised Mean Absolute SHAP importances with the traditional Random Forest (RF) feature importances as shown in Fig.8. The following can be inferred:

- **General Agreement on Top Features:** For both Portrait and Landscape, both RF and SHAP generally agree on *many* of the top-ranked features. Features like `swipe_distance`, `finger_area`, `swipe_duration`, and `pressure` tend to be ranked highly by both methods in their respective orientations. This agreement strengthens the conclusion that these features are truly important for user authentication.
- **Differences in Relative Importance:** While agreeing on *which* features are important, the methods often disagree on their precise relative importance. In Portrait mode, SHAP assigns noticeably higher relative importance to `finger_area` and `y_cord` compared to RF. In Landscape mode, SHAP assigns higher relative importance to swipe distance and finger area compared to RF. RF assigns higher relative importance to `pressure` and `median velocity_last_3_points`.
- **Potential Reasons for Disagreements:** These disagreements can arise because:
  - Traditional RF importance is based on impurity reduction in the trees, while SHAP considers the feature's contribution to the prediction for each instance.
  - SHAP can capture interaction effects between features, which might give higher importance to features involved in complex relationships that RF importance might not fully reflect.
  - The specific structure of the trained Random Forest ensemble can influence RF importance.
- **Orientation-Specific Comparison:**
  - **Portrait:** SHAP seems to emphasize features like `finger_area` and `y_cord` more strongly relative to other features compared to RF.
  - **Landscape:** Both methods see `swipe distance` and `finger_area` as highly important. There are the differences in the ranking of other features like pressure and velocity features.

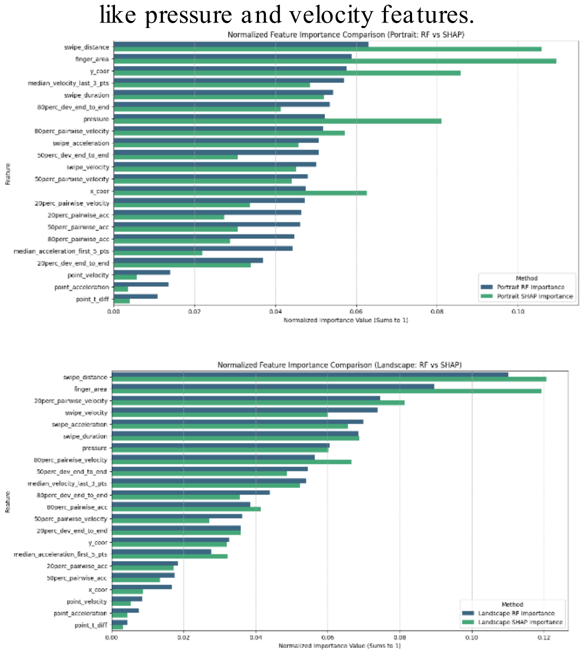


Fig.8. Normalised MASV for Portrait and Landscape Modes

In summary, the comparison shows a good level of agreement between RF and SHAP on the overall most important features, increasing confidence in their relevance. However, SHAP provides a slightly different perspective on the *relative* importance, potentially highlighting features involved in interactions or with non-linear effects more prominently.

IV. PERFORMANCE IMPROVEMENT FROM FEATURE SET PRUNING

Next, we aim to choose representative features from groups of correlated ones, guided by their SHAP importance. We can directly use the insights from the existing heatmaps in Fig. 4 to identify highly correlated feature groups as follows:

- We can see clusters of features with high positive or negative correlations (values close to +1 or -1).

- The percentile features (e.g., pairwise acceleration, pairwise velocity, deviation from end-to-end) are highly correlated with each other.
- swipe_distance and swipe velocity showed a moderate positive correlation, while swipe duration and swipe velocity showed a moderate negative correlation.

A. *Feature Set Pruning Based on a Concensus Mechanism*

From the correlation map in Fig.4, we can select a representative feature from each group using a common correlation threshold of 0.8 to identify highly correlated features. Within each group of features with absolute correlation greater than this threshold, we select the feature with the highest *average* Mean Absolute SHAP importance across both Portrait and Landscape orientations. Features not highly correlated with any others will also be included. This process ensures that we remove features that were *individually* deemed low importance by both RF and SHAP, regardless of whether they were correlated with other features. Even if multiple features in a correlated group have relatively high importance, keeping all of them can introduce redundancy and multicollinearity, which can sometimes affect model stability or interpretability. This step aims to keep only the *most important representative* from each correlated group, ensuring the retained features are both important (by SHAP) and less redundant (by addressing correlation). This mechanism has reduced the feature set from 22 to 18 features.

B. *Model Performance with Reduced Features*

Re-evaluating the Random Forest models on the dataset with reduced features showed that removing these low-importance features did not degrade performance. Instead, the accuracy slightly increased for the portrait model (to approximately 91.34%) and significantly increased for the landscape model (to approximately 90.24%). This demonstrates that the removed features were likely contributing noise or were less informative for the classification task, and their removal improved model efficiency and performance, particularly in the landscape orientation.

TABLE III. CLASSIFICATION ACCURACY COMPARISON

Feature Set	Accuracy-Portrait	Accuracy-Landscape
Original	0.912109	0.829268
Reduced	0.913373	0.902439

V. CONCLUSION AND FURTHER WORK

This work demonstrated the significant impact of phone orientation on the effectiveness of swipe-based user authentication. Through the application of SHAP analysis, we identified the key features influencing user classification in both portrait and landscape orientations, providing valuable insights into the underlying patterns. A notable contribution is the finding that removing features deemed low importance by SHAP analysis can lead to improved Random Forest model accuracy, particularly for landscape swipe data, suggesting that a more focused feature set can enhance performance.

For further work, it would be beneficial to compare the performance of other machine learning models and explore alternative feature selection techniques to potentially achieve even higher accuracies. A crucial next step is to investigate the models' ability to generalise across different devices and orientations, which is essential for real-world applicability. Additionally, incorporating the temporal dynamics of swipe data and validating the findings on larger and more diverse datasets would strengthen the conclusions and broaden the scope of the research.

ACKNOWLEDGMENT

This research has been funded for the project titled, *Exploring the Effectiveness of Explainable Decision Trees for Improving User Trust and Compliance in Continuous Authentication for Dynamic Touch Stroke Biometric Authentication*, May 2023-July 2024, Central QR awards for International Collaboration, Matching grants from the Offices of the Vice Chancellors of the University of Hertfordshire, UK and Multimedia University, Malaysia.

REFERENCES

[1] Evandro S. Ortigossa Thales Gonçalves and Luis Gustavo Nonato "EXplainable Artificial Intelligence (XAI)—From Theory to Methods and Applications," *IEEE Access,* vol. 12, p. 47, 3 June 2024 2024, doi: 10.1109/ACCESS.2024.3409843.

[2] S. Gupta, C. Maple, B. Crispo, K. Raja, A. Yautsiukhin, and F. Martinelli, "A survey of human-computer interaction (HCI) & natural habits-based behavioural biometric modalities for user recognition schemes," *Pattern Recognition,* vol. 139, p. 109453, 2023/07/01/ 2023, https://doi.org/10.1016/j.patcog.2023.109453.

[3] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication," *IEEE Transactions on Information Forensics and Security,* vol. 8, no. 1, pp. 136-148, 2013, doi: 10.1109/TIFS.2012.2225048.

[4] N. Palastanga and R. Soames, *Anatomy and Human Movement: Structure and Function*. Churchill Livingstone, 2012.

[5] M. Martinez-Diaz, J. Fierrez, R. P. Krish, and J. Galbally, "Mobile signature verification: feature robustness and performance comparison," *IET Biometrics,* vol. 3, no. 4, pp. 267-277, 2014, doi: https://doi.org/10.1049/iet-bmt.2013.0081.

[6] A. Serwadda, V. V. Phoha, and Z. Wang, "Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 29 Sept.-2 Oct. 2013 2013, pp. 1-8, doi: 10.1109/BTAS.2013.6712758.

7] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," presented at the Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017.

[8] M. Antal, Z. Bokor, and S. L. Zsolt, "Information revealed from scrolling interactions on mobile devices," *Pattern Recognition Letters,* vol. 56, pp. 7-13, 2015/04/15/ 2015, doi: https://doi.org/10.1016/j.patrec.2015.01.011.

[9] J. Fierrez, A. Pozo, M. Martinez-Diaz, J. Galbally, and A. Morales, "Benchmarking Touchscreen Biometrics for Mobile Authentication," *IEEE Transactions on Information Forensics and Security,* vol. 13, no. 11, pp. 2720-2733, 2018, doi: 10.1109/TIFS.2018.2833042.

[10] U. Mahbub, S. Sarkar, V. M. Patel, and R. Chellappa, "Active user authentication for smartphones: A challenge data set and benchmark results," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 6-9 Sept. 2016 2016, pp. 1-8, doi: 10.1109/BTAS.2016.7791155.

[11] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, "A survey on imbalanced learning: latest research, applications and future directions," *Artificial Intelligence Review,* vol. 57, no. 6, p. 137, 2024/05/09 2024, doi: 10.1007/s10462-024-10759-6.

[12] K. Wang, L. Zhou, D. Zhang, Z. Liu, and J. Lim, "What is More Important for Touch Dynamics based Mobile User Authentication?," *Pacific Asia Conference on Information Systems*, 2020.