Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 7 – No. 36, June 2021 – www.caeaccess.org

Keystroke Dynamics for User-Authentication on Mobile Devices using Ensemble Method

Omoyele Akinsowon
Federal University of Technology, Akure
Department of Computer Science,
P.M.B. 704 Akure, Nigeria

ABSTRACT
As the days go by, the need for security keeps increasing; whether for humans, data, climate and almost everything in existence. However, despite the increase in the need to protect data, information and the corresponding devices, the Personal Identification Number (PIN) is still the most widely used approach on mobile devices. This increased requirement for protection is evidenced by various issues that continue to arise based on intrusion, theft and unlawful access to classified information. It is important to note that apart from secret-knowledge (that is PIN), biometrics and tokens are useful for security even though token seldom authenticates because it is expected to be with the authentic owner. Biometrics on the other hand makes use of the individual himself. This research sought to address the issue of securing data on mobile devices by employing some Machine Learning algorithms. The K-Nearest Neighbours (KNN), Decision Trees (DT) and Multinomial Logistic Regression (MLR) classification algorithms were used to train and test the typing patterns of several individuals who volunteered to type a static passphrase.

After carrying out the different experiments, the predictions given by Decision Trees were the most accurate of all the three base classifiers used with an accuracy value of 99.92%.

General Terms
Security, Behavioural Biometrics, Machine Learning,

Keywords
Keystroke Dynamics, Typing Pattern, Confusion Matrix

1. INTRODUCTION
In today's world, it is rare to find a person who does not use a smart phone, tablet or another mobile device of some sort. Even though brands and service providers vary, almost everyone carries a mobile device today. A report by Google found that smart phone usage rose by 44 percent during the first three months of last year. It also shows that smart phone users are relying more on their devices, with 66 percent of people accessing the Internet daily from a cell phone. Hence, businesses that have a mobile strategy benefit from constant connectivity with their customers since smart phones have become so integral. As of January 2021 there were 4.66 billion active internet users worldwide - 59.5 percent of the global population. Of this total, 92.6 percent (4.32 billion) accessed the internet via mobile devices.In Statista - The Statistics Portal. Retrieved June 29, 2021 from https://www.statista.com/statistics/617136/digital-population-worldwide//#statisticContainer. [1]

GSMA real-time intelligence data shows there are 5.26 billion people that have a mobile device in the world. This means that 67.01% of the world's population has a mobile device. Back in 2017, the number of people with mobile devices was only 53% and breached the 5 billion mark.Statista predicts that by 2023 this number of mobile device users will increase to 7.33 billion.Retrieved March 9, 2021 Source: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world). [2]

According to CNET, there are 2.1 billion Internet users in the world today, which accounts for 30 percent of the world's population. With mobility, people can now access the Internet wherever they are and employees can work remotely wherever there is Internet access. The Google report noted that 80 percent of people do not leave home without their smart phones. When looking up information on a local business, 94 percent use their smart phone while 90 percent of those people follow it up by contacting the business or making a purchase.People who used their smart phone to make a purchase made up 35 percent of people surveyed.

However, the need to ensure security with the use of these devices is very important. This therefore forms a major motivation for this research work. To ensure secure use of these mobile devices, there are several stages ranging from the typing method, capture of individuals' typing patterns, classification of the typing patterns, training the system to recognize the acquired typing dynamics, testing the system to validate its efficiency and finally authentication of users in access control.

The ease of availability of keyboards means that an end user does not typically need to buy any new hardware to use this technology. And this is a big advantage for keystroke dynamics over some other biometric authentication methods, which require specialised hardware such as finger or iris scanners. Another unique advantage gained from incorporating Keystroke Dynamics with the typical password authentication system is that it combines two separate authentication methods into a single input. Typically, when incorporating two-factor authentication, a separate method needs to be introduced in combination with the standard password authentication; such as an authenticator app, which provides a one-time only code. With keystroke biometrics incorporated with a password, the login process remains as straightforward as before, even though the complexity of the verification process has increased.

The fact that no two people have the same typing patterns is a major advantage that this technology has harnessed over the years. Hence, it is hard for an intruder to copy or impersonate because the way a user types is extremely personal to the individual.Different people have varying typing styles and

these form a key aspect of this research work. If it is a standard desktop keyboard for example, an individual will typically use one method of typing, depending on his/her dexterity in typing. For example, a beginner may use the hunt-and-peck method, which involves looking for each key and pressing it one after the other while an intermediate user may use a hybrid typing method, which involves only visually looking for some keys. An expert typist would usually use both hands and type faster than the other two types of users. Hence, there will be different typing styles and dynamics. However, the good thing about this is that users do not regularly switch between these typing methods.

Contrary to standard keyboard, users usually change their typing styles with the use of soft keyboards in mobile devices. In contrast with typing on a standard keyboard, users often change their typing method, depending on their scenario. This switching of typing methods can happen frequently and it poses the question of whether it can be incorporated within a keystroke dynamic system.

There is a variety of features that can be extracted and analysed to create a typing template from a phrase as an individual types; and these features are mainly based on recording the time at which a key is pressed and/or released [4]. A timing vector consists of the keystroke duration times interleaved with the keystroke interval times measured in milliseconds [5]. From the typing dynamics of an individual, there are four separate features that can be produced: Hold time (dwell/down time), flight time (release-to-press time), key press-to-press (key-down to key-down time) and release-to-release time (key-up to key-up time).

The hold time refers to the length of time that a single key is pressed down for. The flight time is the length of time in between releasing a previous key and pressing the next key while the press-to-press time is the time interval from the first key being pressed until the second key is pressed. Similarly, the release-to-release time is the time between releasing the first key and releasing the second key.

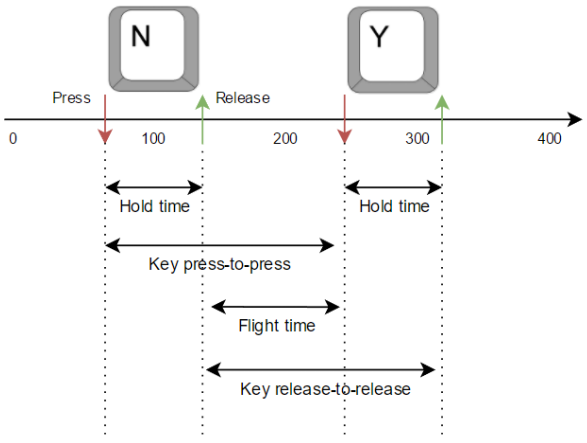

Fig 1 The different keystroke events for characters "N" and "Y" on a timeline

1.2 Related Works

Choi et al [3] used unique keypads that were assigned to and used by only normal users of smartphones to improve the user classification performance capabilities of existing keypads. The experiments were carried out in a controlled environment.

Alsawwan et al [6] studied theuse of keystroke analysis to enhance password security which includes biometrics and typing patterns for everyday users of banking applications. They carried out several experiments and compared the results produced. For example, the authors used the toolkit of WEKA based on method and classifier for the first experiment. In the second experiment, the model suggested by the authors went through analysis with classifiers (such as Naïve Bayes) other than the MLP to see how the model reliability stands against classifications under more than one classifier. At the end of their work, they came to a conclusion that using keystroke dynamics analysis to enhance password security on mobile devices proved to have a great chance of success. One of the limitations of the research was that not enough experiments were carried out. Standard keyboards could have been used in order to bring more successful research results for those keyboards, especially when it is done using other experiment configurations and topologies.

[7] research was motivated by the need of security for the vast use of mobile devices for ease of use and communications because many of the previous studies had focused on Personal Computer keyboard keystrokes. The study investigated the importance of motion features of typing dynamics in addition to the timing features for authenticating individuals and proposed the use of both timing and motion data of the typing patterns of individuals. Various classifiers such as Support vector machine, multilayer perceptron, K-nearest neighbor, and distance-based classifiers were employed to decide the result. At the end of the study, the developed system can be adopted to applications where gender authenticity is crucial, for instance, online dating or online same gender competition exam/game.

[8] showed that the performance of a system that considers both desktop and mobile environments is found to be superior to the best so far. Their work was motivated by the need to have a situation where the performance of Keystroke dynamics is not limited to a particular environment. Hence, a regular keyboard and a mobile device were used in this project. They used the information set concept by computing two types of membership functions (MFs): one based on the timing features of all the samples and another based on the timing features of a single sample. These MFs lead to two types of information components (spatial and temporal) which are concatenated and modified to produce different feature types. The Two Component Information Set (TCIS) proposed for keystroke dynamics-based user authentication improved the performance. The proposed system was not suitable for capturing global characteristics as its main forte is in local characteristics. Also, the choice of membership function limited the overall accuracy because Gaussian function only serves as an effective membership function.

2. SYSTEM CONCEPTUAL MODEL

Keystroke dynamics, keystroke biometrics, typing dynamics and lately typing biometrics, is the detailed timing information, which describes exactly when each key was pressed, and when it was released as a person is typing at a keyboard. User authentication based on typing patterns offers many advantages in the domain of cyber security, including data acquisition without extra hardware requirement, continuous monitoring as the keys are typed, and non-intrusive operation with no interruptions to a user's daily work. Sensor-enhanced keystroke dynamics relies on features derived from keystroke timings and sensor data continuously

sampled while typing. This chapter will explain the entire Keystroke Dynamics system that has been proposed: from data collection to the authentication of users to secure access control on Android devices.

The model of the entire system is a combination of series of steps, phases, techniques and algorithms. At the different stages, several varying techniques are tested and used in order to get effective and efficient answers to the questions that this research seeks to tackle and address. The conceptual models of the system comprise the enrolment model where each user's template is created while the login module handles the decision-making process of confirming if an individual is given access to the device. A score is announced when the user's login template matches the claimed stored template.

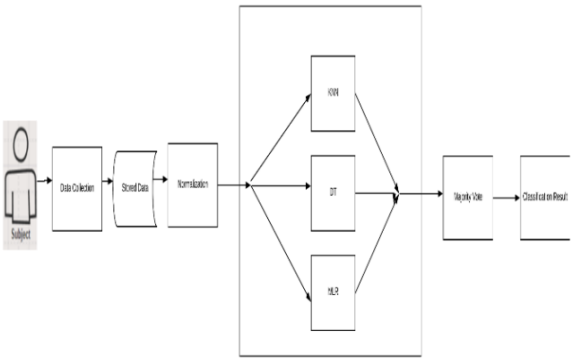

Fig 2 The Ensemble Architecture for the Keystroke Dynamics System

2.1 Data Collection

This first stage of the system architecture is very necessary because the quality of data captured here will determine the overall performance of the system. This is where individual typing data are collected for training, testing and validation of the system. In essence, an Android-based data-collection mobile application was developed. The Android platform was used because of its widespread use and affordability, compared to other mobile platforms.

After deploying the application on the Android device, different people were approached in order to capture their typing dynamics. The data was taken three different times at intervals of 2 weeks apart in order to achieve the aim of this research (that is, to capture peoples' typing patterns when in different states of mind). In fact, each person had to type the password 10 times with the right hand only, the left hand only and then the right and left hands. This method was adopted as a contribution to knowledge because previous works and literature either did not specify which hand was being used to type or considered just one hand at the point of data collection. Hence, each subject typed the password 60 times through the entire data collection stage. As stated earlier, a static password was used- which means that every subject whose typing pattern was collected typed the same password, which was 'comebackaliens'. As they typed these letters, the system captured both their hold timings and flight timings.

The total number of variables or features retrieved at each instance of typing the chosen password phrase (comebackaliens) is expressed with the following notation:

$$N = (2(H + T) - 1) \quad (1)$$

where N, the total number of features for 'comebackaliens' is 27, H the total number of press-to-press timings is 14 and T the total number of hold timings is 13.

As the data is being collected, they are labeled with the name of the subject, the typing method used and the collection session collected so that the data can be split into various ways to suit each of the experiments that were done later. The system was built such that each session of typing dynamics capture is stored separately for every individual.

A data collection application was developed and installed on two different devices. The devices were used to capture the typing dynamics of different individuals at different time intervals. Each person typed the phrase 'comebackaliens' first with the right hand, then the left hand and finally with both hands. The application is comprised of four interfaces where different activities were carried out for the data collection process: Information page, Settings page, Typing Dynamics page and Timings page.

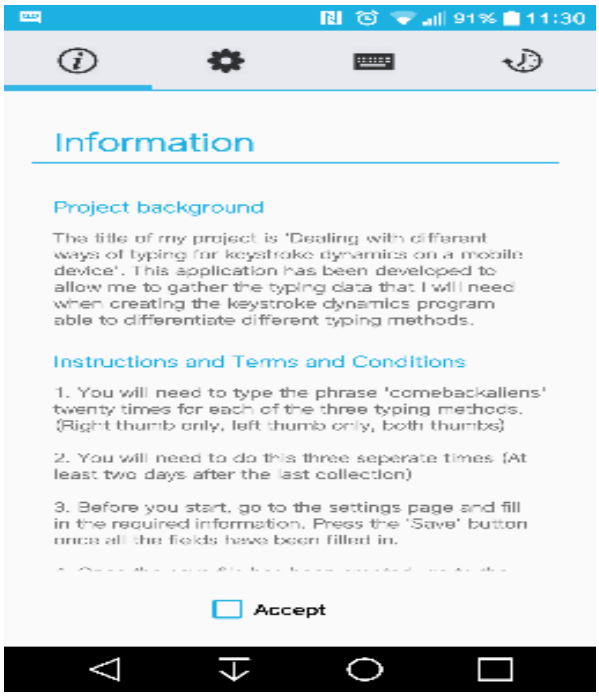

Fig 3 The Information page

2.2 Data Pre-Processing

All the data captured are in terms of timing values in milliseconds and they were stored on the mobile device as notepad files, which were later saved in Excel format for further analysis and fine-tuning for the purpose of further processing.

Data preprocessing involves preparing data in a format that will be suitable to the classification algorithm. This process improves classification performance and also reduces the computational time in training and testing the Machine learning algorithms. The following steps were carried out in the course of processing the keystroke data.

a. Dropping of irrelevant features: Some features that obviously do not contribute significantly to keystroke dynamics' classification were expunged from the data.

b. Non-numeric to Numeric feature conversion: Keystroke data contain mixed feature types (categorical and

numeric features) and the classification algorithm used in this research can only work or perform better on numeric features. Hence, the categorical features were converted to numeric. The categorical feature 'Typing Hand' was converted. The conversion involved assigning unique integer values from zero '0' to each categorical feature value in an alphabetical order. That is mapping categorical features values say $k_1, k_2, k_3, \ldots . k_n$ of a feature category f_i into sequential integer values $, q+1, q+2, \ldots . q+n$. This process is also known as label encoding. A sample of 'Typing Hand' feature and the converted version is shown in the Table 1 and Table 2 respectively.

Table 1. Sample of Categorical Features for Both Hands

SN	TYPING HAND
0	BH
1	BH
2	BH
3	BH
4	BH
5	BH
6	BH
7	BH
8	BH
9	BH

Table 2. Sample of transformed Categorical Features for Both Hands

SN	TYPING HAND
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0

After this phase, the entire data set was split into training set and test set with 70% of data allotted to training the algorithm and the remaining 30% for testing the system performance.

c. Feature Normalization /Scaling: This comes in situations where there is high variation among feature values. This process is necessary to avoid bias problem or issues during classification. In this experiment, min-max normalization was employed to scale high varied feature values into comparable range of zero (0) and one using the equation below

$$v' = \frac{v - min_f}{max_f - min_f} \quad (2)$$

where v' represents the new value, v denotes the observed value (that is, the value to be normalized), max_f and min_f are maximum and minimum values of feature f respectively.

2.3 Data Classification

This stage involves training and testing of the proposed Keystroke Dynamics predictive system with training and test set respectively with 3 different machine learning algorithms – Decision Trees (DT), Multinomial Logistic Regression (MLR), and K-Nearest Neighbor (KNN) and their individual results are combined in an ensemble algorithm in order to maximize the authentication process.

An ensemble algorithm that combined the prediction of three classifiers namely: Decision Trees (DT), Multinomial Logistic Regression (MLR), and K-Nearest Neighbor (KNN) to give a final prediction based on majority voting was adopted.In majority voting, every classifier predicts (votes) one class label, and the highest number of class predicted is selected as the final predicted class label.

The ensemble model performs classification by taking the outputs $V_1(X)$, $V_2(X)$, and $V_3(X)$ of KNN, DT and MLR respectively to make final prediction $C(X)$ using majority voting algorithm. The ensemble starts with K-NN such that given a new or unlabeled keystroke instance, the algorithm will find k (number of neighbours) points in the keystroke training set that are closest to the unlabeled keystroke instance. This is determined using Euclidean distance$d$ between instances in the training set and the unlabeled keystroke instance with features f_i and q_i respectively as follows:

$$d = \sqrt{\sum_{i=1}^{p} (f_i - q_i)^2} \quad (3)$$

where p represents the total number of hand stroke data features. Then, the majority label vote will be selected among classification of the k points as the prediction result $V_1(X)$ of new key stroke instance.

KNN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). The distance between the query-instance and all the training samples of the dataset were calculated and ranked in ascending order to determine the nearest neighbor.

Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree.Decision trees learn from data to approximate a sine curve with a set of if-then-else decision rules. The deeper the tree, the more complex the decision rules and the fitter the model.A decision node has two or more branches. Leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.DT makes prediction $V_2(X)$ by iteratively partitioning the hand stroke training set T into $j$subsets $(t_1, t_2, \ldots, t_j)$ where$j$ is the number of outcome of test over particular feature f_i. The process is continued over each t_k, where $1 \leq k \leq j$, until all elements in each final subset fall under the same class (identity of keystrokes).

Information gain presented was employed to determine the best feature to divide the subset at each stage.

$$IG(T,f) = I(T) - \sum_{v \in values\ (f)} \frac{|T_v|}{N} I(T_v) \quad (4)$$

where v is a value in feature f,$values(f)$ represents all possible values in f, T_v represents instances for which f has v, N represents number of instances in T, $I(T)$and $I(T_v)$ represents entropy of $TandT_v$ respectively.

The MLR pivot will be selected from a set of features. Let yi denote typing features where i=1,2,…,j. Taking note that the probability of a person i being the one who wants to access the device;

$$P(y_1) = \sum(\frac{y_1}{y_i}) \quad (5)$$

The model for any of the features will be

$$log\left\{P\left(\frac{y_1}{y_3}\right)\right\} = \beta_0 + \beta_1 x_i + \cdots \quad (6)$$

$$\text{then, } P\left(\frac{y_1}{y_3}\right) = e^{\beta_0 + \beta_1 x_i + \cdots} \quad (7)$$

where $\beta_0 and \beta_1$are vector weights and x_i the vector of explanatory variables describing feature i while y_3 is chosen as the pivot or base category (any of the other features can be used as the base category).The Majority Vote ensemble method was implemented using Python Programming Language. Each of the classification algorithm produced its predictions, then majority votes were counted, taking the prediction which occurred the most.The results from KNN, DT and MLR were combined to produce a classification that is superior to that of any of the individual algorithms. This is given by

$$C(X) = mode\{V_1(X), V_2(X), V_3(X)\} \quad (8)$$

where $V_1(X)$ is the result from KNN, $V_2(X) is for DT and V_3(X) is for MLR.$

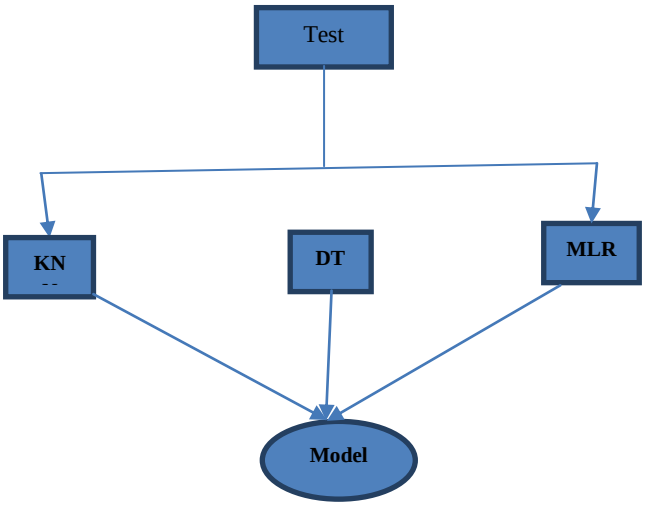

Fig. 4 The Majority Vote Ensemble

3. RESULTS AND ANALYSIS

Each classifier generates a prediction and confidence score, and the prediction with the most "votes" predictions from the ensemble is chosen.The ensemble calculates the mode of the results from the three algorithms used. Afterwards, the evaluation of the proposed model was carried out using standard metrics such as False Acceptance Rate (FAR), False Rejection Rate (FRR) and Equal Error Rate (EER). After various experiments were performed, the result of the base (MLR, DT, and KNN) models and the developed ensemble model were obtained and their performances were compared based on the metrics listed above.

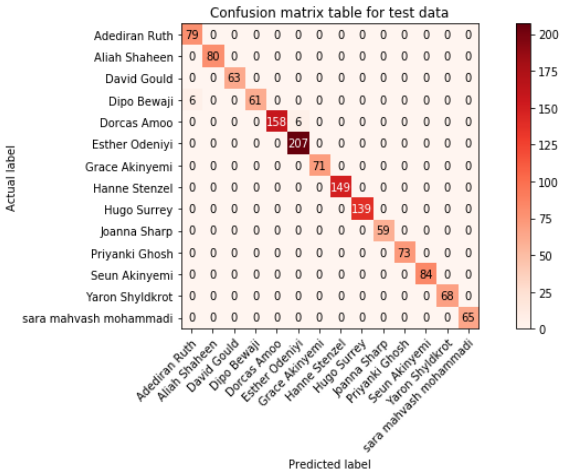

Fig 5. Confusion matrix table on test data for KNN

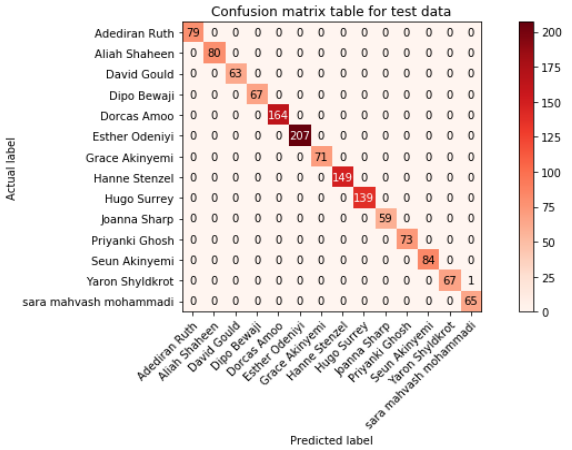

Fig 6 Confusion matrix table on test data for DT

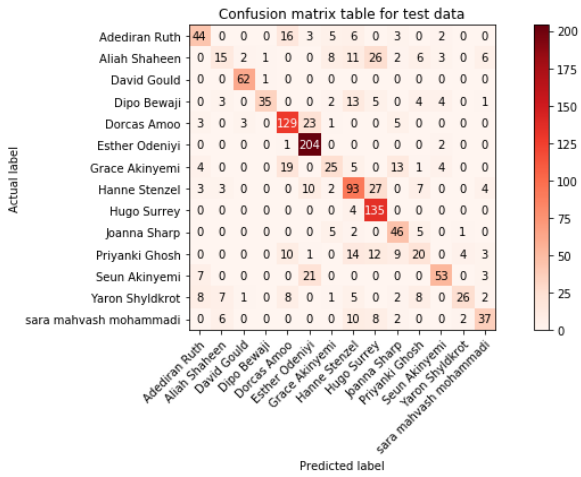

Fig 7 Confusion matrix table on test data for MLR

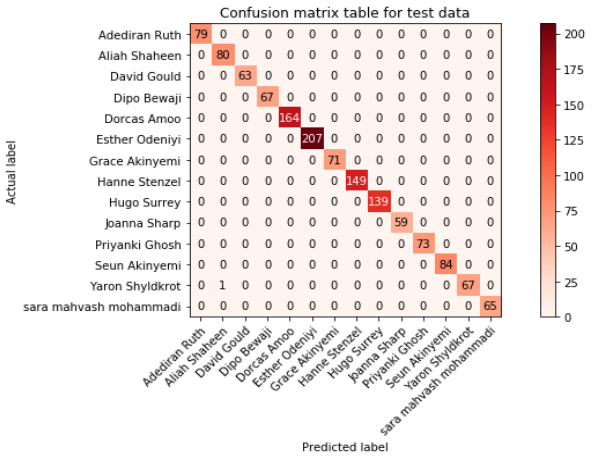

Fig 8 Confusion matrix table on test data for Ensemble Model

The confusion matrices generated from the various experiments that were carried out show that the Ensemble performed better than any of the individual machine learning algorithms. This table 3 gives the details of the results of all the algorithms.

Table 3. Comparison of Majority Vote Ensemble model with base models

Models	Precision	Recall	F1	Accuracy
MLR	66.62	61.44	62.24	67.54
DT	99.86	99.93	99.86	99.92
KNN	99.29	99.07	99.14	99.12
Ensemble	99.92	99.92	99.92	99.92

In the course of carrying out these experiments, the predictions given by Decision Trees were the most accurate of all the three base classifiers used. For instance after evaluation, DT had precision value of 99.92%, recall of 99.92%, F1 of 99.92% and 99.92% accuracy while KNN had precision value of 99.12%, recall of 99.12%, F1 of 99.12% and 99.12% accuracy; and MLR had precision value of 67.59%, recall of 67.54%, F1 of 67.55% and 67.54% accuracy. Afterwards, the Majority Voting Method was used as an ensemble to maximize the effectiveness of the three algorithms. The following values were recorded after evaluating the predictions from theMajority Vote Ensemble: precision value of 99.92%, recall of 99.92%, F1 of 99.92% and 99.92% accuracy. At theend of the research, evaluation revealed that the ensemble enhancedthe individual performances of the base classifiers, thereby affecting the overall system which can be used for authenticating users of mobile devices.

4. CONCLUSION

The main focus of this research work was to develop a system through which users of Android mobile devices can be authenticated in order to secure the different activities that go on with the use of their devices. This goal was motivated by the need to secure peoples' mobile phones which over time has become subtle 'personal assistants'; where they carry out various activities ranging from making/receiving calls and messages, financial transactions, sales and purchases to research and fact-finding. It therefore became imperative to find ways of protecting people from the many crimes that are associated with mobile devices. Android devices were also targeted in this research because of its widespread use and acceptance.

In conclusion, this research has shown that using an ensemble of different algorithms ultimately improved the overall Keystroke authentication system. Different comparative analysis were carried out among the three different classification algorithms and ultimately witth the Ensemble technique. From the analysis, Decision Tree performed best out of the three independent algorithms while K-Nearest Neighbour was next and Multinomial Logistic Regression was the least efficient. This therefore means that DT or KNN can be used alone because in all the evaluated metrics, their individual scores were higher than 90%.

5. RECOMMENDATION AND FUTURE WORK

There is a vast number of Machine Learning classification algorithms that could be used to develop this system. Therefore, further research can be done using some other set of algorithms to see their performance. Also, in the course of collecting keystroke data from the different individuals, the environment and conditions were not exactly the same, even though we tried to make it as a controlled as possible.

6. REFERENCES

[1] https://www.statista.com/statistics/617136/digital-population-worldwide/#statisticContainer

[2] https://www.bankmycell.com/blog/how-many-phones-are-in-the-world

[3] Choi, M.; Lee, S.; Jo, M.; Shin, J. S. Keystroke Dynamics-Based Authentication Using Unique Keypad. *Sensors* 2021, *21*, 2242. https://doi.org/10.3390/s21062242

[4] Banerjee, S. Woodard, D. (2012) Biometric Authentication and Identification using Keystroke Dynamics: A Survey. Journal of Pattern Recognition Research (JPRR), vol. 7, p116-139

[5] Yu Zhong and Yunbin Deng (2015), Recent Advances in User Authentication UsingKeystroke Dynamics Biometrics, DOI: 10.15579/gcsr.vol2.ch2, GCSR Vol. 2, pp. 23-40, 2015, Science Gate Publishing - Available under CC BY-NC 4.0 International License.

[6] AlsawwanAhmed, AlrashdanMaen, Al-MaatoukQusay, AbdulnabiMohamed, Veeramani, IndrianVijai, TubishatMohammad, Mosab Tayseer AlRashdan, Al-Othman Abdulaleem Z (2020), Keystroke Dynamics Analysis to Enhance Password Security of Mobile Banking Applications,

[7] Lee Hyungu, Yeon Hwang Jung, Kim Dong In, Lee Shincheol, Lee Sung-Hoon and Shin Ji Sun (2018), Security and Communication Networks, Hindawi, Volume, Article ID 2567463, 10 pages, https://doi.org/10.1155/2018/2567463

[8] Bhatia, Aparna & Hanmandlu Madasu (2017), Keystroke Dynamics Based Authentication Using Information Sets. Journal of Modern Physics. 08. 1557-1583. 10.4236/jmp.2017.89094.