Intelligent Automation And Soft Computing, 2019
Copyright © 2019, TSI® Press
Vol. 25, no. 4, 651–661
https://doi.org/10.31209/2018.100000060

Feature Selection and Representation of Evolutionary Algorithm on Keystroke Dynamics

Purvashi Baynath, Sunjiv Soyjaudah, Maleika Heenaye-Mamode Khan

University of Mauritius, Reduit, Mauritius

ABSTRACT

The goal of this paper is (i) adopt fusion of features (ii) determine the best method of feature selection technique among ant Colony optimisation, artificial bee colony optimisation and genetic algorithm. The experimental results reported that ant colony Optimisation is a promising techniques as feature selection on Keystroke Dynamics as it outperforms in terms of recognition rate for our inbuilt database where the distance between the keys has been considered for the password derivation with recognition rate 97.85%. Finally the results have shown that a small improvement is obtained by fused features, which suggest that an effective fusion is necessary.

KEY WORDS: Artificial Bee Colony Optimization, Ant colony Optimization, Artificial Neural Network, Biometrics, Genetic Algorithm, Keystroke Dynamics.

1 INTRODUCTION

WITH the global access to information and resources, the amount of data that is accessed and transmitted on a daily basis using the internet is phenomenal. Nowadays people have the possibility to access their personal, social and financial information through the internet everywhere. The combination of username and password has offered users an authentication system to login to the computer system. This process usually ensures that the users have access to their own information. However, its simplicity has made it vulnerable to shoulder surfing, guessing and dictionary attacks (Shanmugapriya and Padmavathi, 2009).

To strengthen the password, the latter is combined with biometric technology. Biometric techniques have proved to be an excellent verification mechanism, being hardly compromised since it cannot be stolen or lost and it is unique for every individual. Keystroke dynamics is one of the biometric solutions which can solve the problem of sending secure information through internet. It analyses the way a user types at a terminal. By simply monitoring the users' typing rhythm pattern, the identity of a person can be determined (Shanmugapriya and Padmavathi, 2009). Keystroke dynamics is a behavioral biometric approach which is cheap and does not need any sophisticated hardware other than a keyboard. This criterion has made it easily acceptable by user (Giot, Dorizzi and Rosenberger, 2011).

A variety of natural and biological processes have been the motivating factor in the creation of nature inspired algorithms. These algorithms have gained popularity due to the biological systems to efficiently adapt to the frequently changed environments. Some algorithms and concepts that have been motivated by nature are evolutionary computation, neural networks, ant colony optimization, particle swarm optimization, artificial immune systems and bacteria foraging algorithm among others (Shanmugapriya and Padmavathi, 2009, Giot, Dorizzi and Rosenberger, 2011). Swarm behavior encompasses the study of different colonies of social insects like bees, ants and termites. This type of behavior can be categorized by autonomy, distributed functioning and self-organization. In the last two decades, researchers have been studying different social behavior insects in an attempt to use swarm intelligence concepts and build various artificial system.

Until now, Ant colony optimization (ACO), artificial bee colony optimization (ABCO) and genetic algorithm (GA) have not been fully explored and applied on the same features of Keystroke Dynamics as a feature selection technique. As a novelty, the flight time and dwell time of each feature have been fused. The application of the above mentioned feature subset selection is then applied and explored on the

CONTACT Purvashi Baynath p.baynath@gmail.com

fused features of keystroke dynamics. The paper is structured as follows; Section 2 covers the literature review of the work carried on keystroke dynamics. The comparison on the data collection techniques as well as the feature extraction has been detailed in section 3. Section 4 shows the results of the simulation and Section 5 evaluates the results.

2 RELATED WORK

KEYSTROKE dynamics (KD) is considered as a strong behavioral biometric based authentication system (Shanmugapriya and Padmavathi, 2009). Like any biometric system, Keystroke dynamics has four important modules known as sensor module, feature extraction module, matching module and decision module (Ross and Jain, 2004). From the studies conducted, it is clearly indicated that the crucial phases within a biometric system remain the pre-processing, processing and matching (Monrose and Rubin 1997, O'Gorman, 2003).

The statistical technique of keystroke dynamics approach has been commonly studied by researchers (Monrose and Rubin 1997). By applying statistical technique, the performance of proposed algorithm can be easily analysed compared to another algorithm. Hence in this work, the statistical method has been used. As previously stated, the feature extraction technique is a crucial part of any Biometric system. Initially, the capable features of KD were the flight time and press time of each key (Karnan, Akila and Krishnaraj, 2011, Teh, Teoh and Yue, 2013, Syed and Syed, 2015), then Davoudi and Kabir(2009) propose the distance between the typing sample as a possible feature. The idea of considering the distance between the typing samples has been motivating and hence for our data capture the distance between the keys has been taken into consideration.

The feature that has been commonly used in KD system is the flight time and dwell time. The work that has been conducted using flight time and dwell time is detailed below. Monrose et al.(2000) used the flight time between fixed texts and applied euclidean distance and Bayesian classifier define the variability with which typists produce digraphs and yield 16.78% of FRR and 7.83% of FAR. Cho et al. (2000) used the delay of Dwell time and then processed in a multilayer perception neural network so as to discriminate between the user and an imposter. By adjusting the threshold, the results achieved 0.0 % of FAR and approximately 1 % of FRR. On the other hand, Rajput et Vijayawargiya (2011) extracted both flight time and dwell time in order to study the emotional state of a user. Shanmugapriya and Padmavathi (2011), applied both flight time and dwell time to determine the best algorithm to be applied to increase security. Teh et al. (2012) have used dwell time and flight time and the results of the study demonstrated that the finest performance is obtained after applying dwell time and flight time. Obaidat and Sadoun (1997) studied the digraph latencies and key hold times using multiple machine learning algorithms.

In this section, the inspiring work which has been conducted using the nature inspiring algorithm is detailed. Karnan et al (2009) has applied the nature inspired algorithm namely Particle Swarm Optimisation (PSO), Genetic algorithm (GA) and Ant Colony Optimization (ACO) for its feature subset selection module on Keystroke dynamics features [30]. The result obtained was of 92.8% classification accuracy in favor of Ant colony optimization when compared to PSO and GA. On the other hand, Shanmugapriya and Padmavathi (2011) have also applied the ACO, PSO and GA algorithm with Extreme Leaning Machine for feature subset selection where the results gain was of 46.51 % feature reduction compared to GA and PSO. Senapathi and Batri (2014) proposed a new Genetic Algorithm wrapper approach. From the analysis made, a new and improved version called PSO method has been observed compared to Genetic Algorithm (2014). Another inspiring work that has been conducted by Nisha and Kumar where they have proposed an enhancing control user authentication on username and password. In their work, the authors have used GA, PSO and ACO as the feature subset selection. Even if the results gained during the experiments conducted above were impressive, however much details of the samples used were not communicated.

The data capture plays a vital role for KD system. It is to be noted that the several online database for KD is already available. The Table 1 shows some of the available databases. Even if these databases are already available online, the different protocols for data acquisition have not been communicated. In any biometric system, the environment of the data capture plays a very important role as the keystroke depend on various factor like the typing position of the user, the hand used for the data capture, the climatic condition, and much more. These databases differ from each other by the number of individuals, separation between sessions, the acknowledgement of the password, the used keyboards, (which may deeply influences the way of typing), and the use of different or identical passwords, (which impacts on the quality of impostors' data) (Killourhy and Maxion, 2009).

Table 1 Online Available database for keystroke system

Database	Feature	Number of user	Input/keyboard	Text Type
Killourhy and Maxion	Timing	51	.tie5Roanl / QWERTY	static
Giot et al.	Timing	133	Greyc Laboratory	static
Allen	Timing	104	Jeffrey, Allen, drizzle	Static

For our data capture process in our inbuilt database all the above mentioned condition has been taken into consideration and the distance of the keys position on the keyboard has also been taken for the password derivation. Inspired from the previous work conducted using flight time and dwell time of each digraph, in the study, as a novelty the fusion of the flight time and the dwell time of each keys are applied on evolutionary algorithm to deal with the challenges faced in keystroke dynamics. Ant Colony Optimization, Artificial Bee Colony Optimization and Genetic Algorithm have also been applied on the fusion of flight time and dwell time for two different databases where our inbuilt database has taken the distance between keys for the password derivation.

3 APPROACH

KEYSTROKE dynamics evaluations involve the following steps:

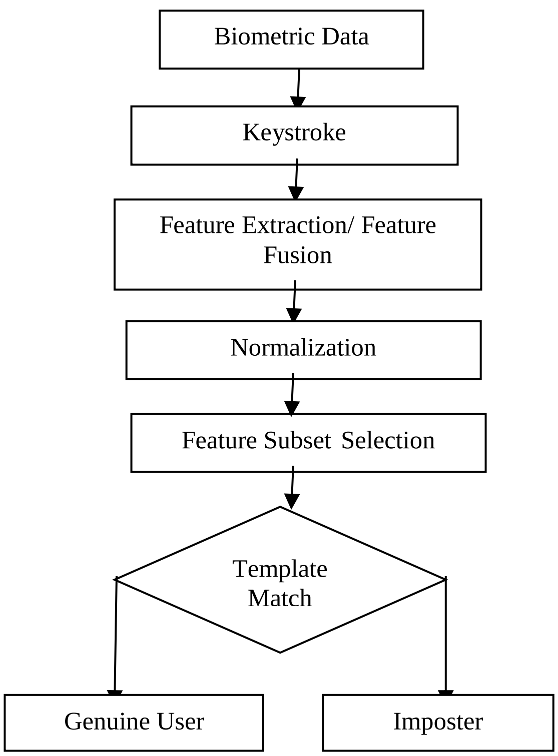

Figure 1: Methodology to develop the keystroke Dynamics

3.1 Recruiting subjects and data collection

A database was created with subjects from University of Mauritius. The data collection was carried out from 1000 users each providing 2000 typing samples for each password. The passwords used for data collection are: .tie5Roalnb and aeR5t.ilnb. For the password derivation process, the distance of keys placement on the keyboard was considered. During the data capture process, two types of conditions were respected. First the users were asked to type the password provided to them using both hands and then for the second set of data captures, the users were requested to type in the given password using only one hand (their strong hand). For the data capture, Baynath et al. (2016) have adopted a new way for collecting data from the user. Passwords were devised in such a way that the distance of the keys placement on a keyboard is distant. For data capture, the controlled and static environment is optimum for system configuration. The data captured were done in a laboratory at the University of Mauritius. The laboratory was well ventilated so that the user does not feel restless while the keying process was executed. All the precautions were taken to have a constant environment for all the users. In keystroke dynamics, the typing position (sitting, leaning or standing) affect the captured data during enrolment and authentication. A desired typing position, that is, sitting with a straight back produces a more accurate representation of the users characteristics. The seating positions of the users were further adjusted so that glare and the lighting conditions do not affect them. The data capture was done on a laptop with an external QWERTY keyboard connected to it. The laptop was chosen for the data capture so that the data can be stored in only one database. The external keyboard was connected as from the ergonomics researches conducted; it was found that the integrated keyboard on laptop makes the users turn their wrists in order to type on the keyboard which is not comfortable for the user. An external screen was also connected so that the subject does not have to bend their neck down to look at the screen. An adjustable typist chair was used, so that each user is in a comfortable zone for the data capture. To authenticate the subject correctly, the data capture of the user must be more or less the same for each digraph each time the user key in. Environmental condition was carefully maintained for the data capture of the user. The database was devised in line for the evaluation of the proposed techniques. The Euclidean distance formula has been used to determine the distance between the keys.

In this research, the online database choice is Killourhy and Maxion (2009). The password used by the latter is '.tie5Roaln' which resemble the password convention adopted for our password derivation. Moreover, this online database contains the dwell time as well as the flight time of the digraph of keys, which has been used for the fusion process.

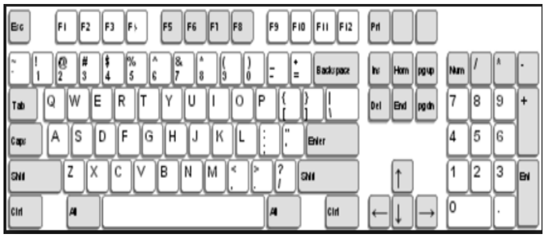

Figure 2: **Types of Keyboard used.**

3.2 Recording the timing information

The second step which has been devised is the recording of the timing information. There are different ways to capture the timing information; the flight time or dwell time can be adopted. For this study, an application has been designed to capture the flight time, dwell time and the distance between the keys of each key for the passwords .tie5Roalnb and aeR5t.ilnb. The recording has been made for the digraphs. Digraph comes under Press to press category. Digraphs contain two consecutive keystrokes. As example do demonstrate the above said is “system”, the digraph is (‘sy’, ‘ys’, ‘st’, ‘te’, ‘em’).

3.3 Pre-processing

Extracted features contain much unnecessary information like noise. Normalization techniques were adopted to eliminate the unwanted impurities. For this purpose, among the normalization techniques that exist, Z-score normalization has been used. Z-score normalization has been chosen as it is robust and has a high efficiency compared to other normalization techniques.

The pre-processed results are then given to the next step namely the feature subset selection.

3.4 Feature subset selection

Feature subset selection is applied to high dimensional data before moving to the next step that is the classification process. Feature subset selection is a process that selects a subset of original features. Feature subset selection reduces the number of features, removes the irrelevant, redundant, noisy data and speeds up the results obtained from various algorithms. So far Ant Colony Optimization (ACO), Artificial Bee colony optimization (ABCO) and Genetic algorithm (GA) which are promising techniques has not been fully developed for feature subset selection on dwell time and flight time of keystroke dynamics (Senavati and Bari, 2014). To address the challenges and shortcomings presented in existing methods used to select feature for keystroke dynamics, the Ant Colony Optimization (ACO), Artificial Bee colony optimization (ABCO) and Genetic algorithm (GA) are being explored. For the techniques used the original algorithm for ACO, ABCO and GA has been further explored and applied on the fusion of flight time and dwell time features of two different databases in this paper. The strong hand concept has been used in order to consider the distance between the keys as while typing using both hands, it would have been difficult to determine if ever the distance affect the performance of system.

3.4.1 Ant Colony Optimization (ACO)

The ant colony optimization technique has been inspired by the examination on real ant colony’s rummaging activities for food. These ants can frequently recognize the shortest path between food source and their nest. Ants broadcast information with the assistance of volatile chemical substances that they left in their crossing way and also called as the “pheromone”. By this process the goal of identifying the shortest path to identify food sources is achieved. An ant recognizing an already laid trail can identify the thickness of pheromone trail. It chooses with high probability a shortest path and strengthens that trail with its own pheromone (Senavati and Bari, 2014).

In this work, the mean and standard deviation have been computed for each digraph of flight time and dwell time present for each sample. After, the application of z-score normalization, the two features (flight time and dwell time) has been fused. Then, Ant colony algorithm has been used for selecting the optimum feature for each participant and the selected features are considered for future classification. The steps use for the simulation is provided below:

Step 1. The feature value a[y] was determined after the application of normalization for the flight time and dwell time of each digraph of keystrokes. Then the fusion of these flight time and dwell time features was performed.

Step 2. The fitness function F[y] was calculated by the following equation for every F[y].

$$F[y] = 1 / (1 + a[y]) \qquad (1)$$

Step 3. The following criterion has been initialized as shown below:

Number of iterations, NI = 5

Number of Ants, NA = 2

Initial pheromone value for every a[y], T0 = 0.001

Rate of pheromone evaporation parameter for every a[y] , A = 0.9

Step 4. The fitness function values was stored in S, where S = {F[y], T0, flag} where flag column mentions whether the feature is selected by the ant or not.

Step 5. The following was repeated for NI times:

A random feature value g[y] in a[y] is selected for each ant with the criteria that the particular feature value should not have been selected previously.

Selected feature value’s, pheromone value is updated by the following:

$$Tnew = (1 - \alpha)\ y\ Told + \alpha\ y\ Told \text{ for } g[y] \qquad (2)$$

Where Tnew and Told are the new and old pheromone value of the featured value.

The Lmin is obtained by equating it to min (g[y]) where Lmin is the Local minimum. Initially the fitness value a[y] is directly assigned to Local Minimum (Lmin). Then the next fitness value a[y] is compared with the previous value already calculated. The minimum is found in them and is replaced with the Local minimum value.

If Lmin <= Gmin then Gmin = Lmin is assigned. Else no change in Gmin value where Gmin is the Global minimum.

The best feature a[y] is selected, whose solution is equal to the Local minimum value at the end of the last iteration.

The selected g[y]'s pheromone value is globally updated

g[y],

$$\text{Tnew} = (1 - \alpha)\ y\ \text{Told} + \alpha\ y\ \Delta\text{Told} \quad \text{for } g[y] \quad (3)$$

Where α is a rate of pheromone evaporation parameter,

$$\Delta = 1 / \text{Gmin}$$

The remaining ants and their pheromone are updated as:

$$\text{Tnew} = (1 - \alpha)\ y\ \text{Told} \quad (4)$$

Where α is a rate of pheromone evaporation parameter.

Finally, the Gmin value is stored as optimum value.

These steps have been applied on the fused features of flight time and dwell time for the two different databases.

3.4.2 Artificial Bee Colony Optimization (ABCO)

Another technique for mimicking the environment is the Artificial Bee Colony Optimization (ACBO) technique. The characteristics of the mentioned technique are justified for its application on keystroke dynamics features. Bee colony optimization is inspired from the activity of honey bee exploring the environmental in search of flower patches (nectar) and indicates the food source to the other bees of the colony when they returned to their hive. The steps use for the simulation is provided below:

Step 1. The food source positions are initialized.

$$\bar{\upsilon}\text{i} = (\text{i} = 1, ..., \text{SN})$$

Solutions are randomly produced in the range of parameters where SN is the number of the food sources.

Step 2. Each employed bee produces a new food source in her food source site and exploits the better source.

For each employed bee, whose total number equals to the half of the number of food sources, a new source is produced by (5):

$$\text{Vij} = \text{yij} + \varphi\text{ij}\ (\text{yij} - \text{y kj}) \quad (5)$$

Where φij is a uniformly distributed real random number within the range [-1, 1], k is the index of the solution chosen randomly from the colony

$$(\text{K} = \text{int}\ (\text{rand} * \text{SN}) + 1\ ) \quad (6)$$

And j = 1,….. D where,

D is the dimension of the problem. After producing ῡi, this new solution is compared to ῡi solution and the employed bee exploits the better source.

Step 3. Each onlooker bee selects a source depending on the quality of her solution. A new food source is produced in the selected food source site and the better source is exploited. An onlooker bee chooses a food source with the probability (8) and produces a new source in selected food source site by (7). For employed bee, the better source is then decided for the exploitation.

$$\text{Pi} = \text{fiti} / \Sigma\text{SNj=1 fitj} \quad (7)$$

Where i fit is the fitness of the solution i x r.

Step 4. The source to be abandoned is determined and its employed bee as scout is allocated for searching new food sources.

After all onlookers are distributed to the sources, they are checked whether they are to be abandoned. If the number of cycles that a source cannot be improved is greater than a predetermined limit, the source is considered to be exhausted. The employed bee associated with the exhausted source becomes a scout and makes a random search in problem domain by (8).

$$\text{Vij} = \text{Vjmin} + (\text{Vjmax} - \text{vjmin})\text{* r and} \quad (8)$$

Step 5. The best food source found is memorized so far.

Step 6. The steps 2-5 have been repeated until the stopping criterion is met.

These steps were followed for the application on the fusion of flight time and dwell time for the two different databases.

3.4.3 Genetic Algorithm (GA)

A genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. This method is thus simulated on the fusion of flight time and dwell time features of keystroke dynamics.

In genetic algorithm, the processes applied are as follows (Shanmugapriya and Padmavathi, 2009):

Step 1. The number of chromosomes, generation, and mutation rate and crossover rate value are determined.

The number of chromosomes in population is defined, and then a random value is generated for the other chromosomes.

Step 2. Chromosome-chromosome number of the population is generated, and then the initialization value of the genes chromosome-chromosome is made with a random value.

Step 3. Steps 4-7 are processed until the number of generations is met

Step 4. The fitness value of chromosomes is evaluated by calculating objective function using the function fy where, the evaluation step is carried out, where a, b, c, d are randomly selected chromosome.

$$fy = ((a + 2b + 3c + 4d) - 30) \quad (9)$$

Step 5. Chromosomes selection

The selection is made using the formula below:

$$Fitness\ [1] = 1 / [\ 1 + F\_obj\ [1] \quad (10)$$

The probability for each chromosome is formulated by:

$$P[i] = Fitness\ [1] / Total \quad (11)$$

Step 6. Crossover

Pseudo-code for the crossover process is as follows: begin k← 0; while (k< ρc) then select Chromosome [k] as parent; end; k = k + 1; end; end;

Step 7. Mutation

Total gen = number of gen in chromosome * number of population (12)

Step 8. New Chromosomes (Offspring)

Step 9. Solution (Best Chromosomes)

After the application of GA on the fused features of the flight time and dwell time of keystroke dynamics, the appropriate classification technique is being adopted to determine the best feature subset selection among them.

3.5 Classification

For the techniques adopted for feature subset selection namely Ant colony optimization, Artificial Bee Colony Optimization (ABCO) and Genetic Algorithm, artificial neural network would be appropriate to be used for the classification phase. In this phase, the algorithms used would search whether the test set matches any templates from the database.

The artificial neural network works just like human brain. During the learning process (training) or when the system is being operated (after being trained). The feedforward network has been used. The pattern of the fusion of flight time and dwell time of each database has been fed into the network via the input units. Then the hidden unit is triggered and then the desired output is obtained from the output unit. In the system, each units trigger at its own time. The individual unit receives the input from the other units to its left and then these units are multiplied by the weights of the connection along which they are travelled.

The summing up of every unit is made and then when this sum is greater than a certain threshold, then the next unit gets triggered (the unit present on the right side). When the network has been trained with enough learning examples, then a new sample can be feed to the system and its behavior is analysed. The parameter that has been used is detailed in Table 2.

Table 2: Parameters used for Neural Network

Input data size	100000
ANN Type	MLP with N hidden layers
ANN training Method	Backpropagation with Levenberg and Marquardt Optimization
Average training epochs	MLP 100 to 3000
Mean Square Error MSE goal	10-4

The biometric matching has helped in comparing the biometric templates to determine the degree of similarity. The matching score was compared to the threshold value. When the match score was exceeding the threshold value, the result was a match else it was miss-match.

The algorithm has calculated the difference between the trained values with the tested value for each password used. The threshold value has been set at random. Using the threshold value, it has been determined if the person is accepted or rejected. From the results obtained the False acceptance rate (FAR) and False rejection rate (FRR) has been determined. The data has passed through the neural network after the feature selection process. The output obtain were calculated in terms of FAR (false acceptance rate) and FRR (false rejection rate) for each feature subset selection techniques used

The proposed system has made use of previously captured data of dwell and fight time of different password from our inbuilt database and killourhy and Maxion database [33].

4 EXPERIMENTAL RESULTS AND DISCUSSION

EXPERIMENTS were conducted in order to validate the proposed techniques namely ACO, ABCO and GA. Initially, it is very important in this type of experiment to set the threshold value. The value will eventually differentiate between genuine users and imposters. To achieve this, a training set of 20 subjects each having 300 password samples were used. 200 samples were used for the training set of the neural network and 100 password samples were used for the test set. Note that the same instance of the user was used in the test set and training set. Table 2 summarizes the result obtained with different threshold value.

For evaluation purpose, we often ask the question if ever different datasets can be used to compare the evaluated results. The evaluation of a distinct proposed technique of KD with one or more dataset is not common. However to check the reliability of the proposed technique, it becomes important to test the proposed system with different database.

The threshold is set depending on the security that the user wants to have for their application. From the experiments conducted, it is concluded that threshold value 0.7 provides a better overall performance with recognition rate. (RR) 96.55%, false acceptance rate (FAR) of 0.12 and False rejection rate (FRR) of 1.50. It is to be noted that the experiments were conducted 20 times and the average were taken. If a larger margin of values of threshold were considered, the security of the system was being defeated by generating a FAR of 2.2% though the recognition rate is higher and with a lower value of threshold also the performance deteriorated.

Table 4 provides an overview of training time and testing time required for both databases using each of the feature subset selection technique.

After setting the threshold value, several experiments were conducted to evaluate the performance of the system by applying ACO, ABCO and GA. As explained in previous section, a fusion of flight time and dwell time were considered. The algorithm has been tested using (1) the online database (2) inbuilt database where the user was asked to type with only one hand so that the distance between the keys can be considered (3) inbuilt database where the user was asked to key in using both hands. The two types of data capture have been done as the distance has been considered for the data capture for our inbuilt database. First the data were normalized using z-score normalization technique. The fusions of dwell time and flight time of the different passwords have been applied on each feature subset selection technique namely ACO, ABCO and GA. The threshold value was set to 0.7 throughout the experiment. Table 5 summarizes the result obtained for the application of feature subset selection in our inbuilt database. The parameters has been used is detailed below.

Table 3: **Threshold value for keystroke system**

Threshold value set	Recognition Rate (RR) %	False Acceptance Rate (FAR) %	False Rejection Rate (FRR) %
1.5	79.40	8.00	3.60
1.3	89.40	4.20	2.60
0.7	96.55	0.12	1.50
0.8	97.36	2.20	0.44

Table 4 **Training Time and testing time required for the database**

Feature Selection	Inbuilt database		Killourhy Database	
	Training (ms)	Testing (ms)	Training (ms)	Testing (ms)
Ant colony Optimization	20	0.70	25	0.8
Artificial Bee Colony Optimization	36	0.92	39	0.85
Genetic Algorithm	35	0.90	34	0.89

The experiments were repeated for the same users on different feature subset selection technique. A recognition rate of 97.85% was obtained for Ant Colony Optimization compared to Artificial Bee colony Optimization with recognition rate of 94.52% and genetic algorithm with a rate 93.50% for 1000 subjects when the user was asked to use their strong hand. A recognition rate of 92.2% was obtained in favor of ACO when compared to BCO and GA when the user was asked to use both their hands.

The same experiments were conducted using the online database. Table 6 shows the result obtained.

The same sets of data were used for the experiments on different feature subset selection technique for the Killourhy and Maxion database. A recognition rate of 71.50% was obtained for Ant Colony Optimization compared to Artificial Bee colony Optimization with recognition rate of 72.52% and genetic algorithm with a rate 79.00% for 1000 subjects. It is to be noted that the hand that has been used for the data capture has not been communicated of the Killourhy and Maxion database.

Figure 3 shows the graphical representation of the results obtained for the recognition rate for both the database used. The application of different dataset has shown the behavior of different algorithm behave the same trend.

Several authors have proposed techniques based on the evolutionary algorithm. Table 7 shows a comparison of our experiment with some of the existing technique which gained remarkable results. It can be observed that the overall performances of all the techniques applied in the experiments have achieved more that 60% recognition rate.

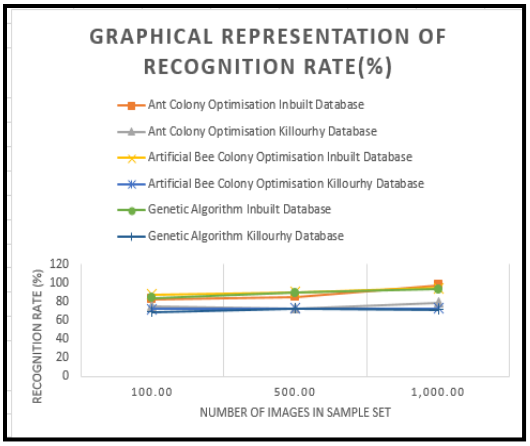

Figure 3: **Graphical representation of results.**

Karnan et al, has applied the ACO technique as the feature subset selection to obtain a recognition rate of 64.88% with KNN classifier whereas Shanmugapriya and Padmavathi (2011) has made the application of GA, Swarm optimization and ACO to reduce the samples by using Extreme Learning Machine. Nisha and Kumar (2014) have applied the same evolutionary algorithm as applied by us in the experiment except that the renovated artificial Bee Colony optimization was used whereas in our work Artificial Bee Colony Optimization has been applied. However, the number of samples used for the simulation was not communicated by Nisha and Kumar even if the results were remarkable. The latter have achieved the best result using the renovated artificial Bee colony optimization as the feature subset selection technique whereas in our experiment, the best result has been achieved using Ant Colony Optimization. However, the results achieved in our experiment have yielded much higher recognition rate compared to Nisha and Kumar.

Comparing the results obtained during our experiment using the database of Killourhy and Maxion and our inbuilt database, the Ant colony optimization is better in terms of feature selection technique. However our database which contains data captured for passwords with more distance between the keys gained a higher recognition rate compared to Killourhy and Maxion database. From the results of the experiment conducted and the shape of the graph in Figure 3, it is observed that Ant colony optimization is a better in terms of feature subset selection compared to the other techniques even if different dataset has been used to compare the results.

5 CONCLUSION

SECURITY and authentication are of great concern in computer networks or systems. Keystroke dynamics is a cheap biometric system that is used to improve security. In order to improve security further, an authentication system using Keystroke features flight time and dwell time has been developed. In this research work, evolutionary algorithms were applied on Keystroke Dynamics fused Features. A customized database was developed where data was collected from 1000 subjects. Note that environmental factors were taken into consideration during the experiment. ACO, ABCO and GA were then applied on the fused features of flight time and dwell time feature to extract the subset of original features. Several experiments were conducted where the RR, FRR and FAR were determined. It was concluded that ACO is a promising feature subset selection technique since it has achieved a RR of 97.85%. The higher recognition rate has been achieved when the distance between the keys is higher. Hence it is advised to use a password with more distance between the keys position on the keyboard while choosing a password. Compared to techniques elaborated in literature our methods have achieved some improved results.

Table 5: **Results obtained for keystroke dynamics on our inbuilt database**

Feature Selection	Number of images in sample set	Recognition Rate (RR) %		False Acceptance Rate (FAR) %		False Rejection Rate (FRR) %	
		Data collected with one hand	Data collected with both hand	Data collected with one hand	Data collected with both hand	Data collected with one hand	Data collected with both hand
Ant colony Optimization	100	83.00	80.0	2.00	12.0	9.00	8.00
	500	85.00	81.0	0.85	8.00	14.15	11.00
	1000	97.85	95.2	0.15	2.20	2.00	2.60
Artificial Bee Colony Optimization	100	87.52	86.1	4.00	2.66	8.50	9.82
	500	90.48	85.2	3.00	5.24	6.52	9.56
	1000	94.52	91.3	2.00	4.18	3.48	4.52
Genetic Algorithm	100	84.20	79.2	5.90	9.10	9.90	11.70
	500	89.52	83.4	3.00	4.20	7.48	12.40
	1000	93.50	91.6	1.50	4.15	5.00	4.25

Table 6 Results obtained for online database of keystroke dynamics

Feature Selection	Number of images in sample set	Recognition Rate (RR) %	False Acceptance Rate (FAR) %	False Rejection Rate (FRR) %
Ant colony Optimization	100	75.00	6.00	19.00
	500	73.00	4.50	22.50
	1000	79.00	3.50	17.50
Artificial Bee Colony Optimization	100	72.00	5.20	22.80
	500	73.00	6.50	20.50
	1000	72.52	2.90	24.58
Genetic Algorithm	100	69.20	11.20	19.60
	500	72.59	8.50	18.91
	1000	71.50	9.50	19.00

Table 7: Comparison of Feature subset selection

Author	Classifier	Technique	Recognition Rate (RR)
Karnan et al (2009)	KNN	Ant Colony Optimization	64.88
Nisha and Kumar(2014)	Back Propagation Neural Network	Swarm Optimization	89.23
		Genetic Algorithm	87.54
		Ant Colony Optimization	92.80
		Renovated Artificial Bee Colony Optimization	93.50
Shanmugapriya and Padmavathi (2011)	Extreme Learning Machine	Genetic Algorithm	Features reduced by 34.880
		Swarm Optimization	Features reduced by 30.230
		Ant Colony Optimization	Features reduced by 46.510
Our experiment (Our database)	Back Propagation Neural Network	Ant Colony Optimization	97.85
		Artificial Bee Colony Optimization	94.52
		Genetic Algorithm	93.50
Author	Classifier	Technique	Recognition Rate (RR)
Our experiment (Killourhy database)	Back Propagation Neural Network	Ant Colony Optimization	79.00
		Artificial Bee Colony Optimization	73.00
		Genetic Algorithm	72.59

6 REFERENCES:

P. Baynath, K.M.S Soyjaudah and M. Heenaye-Momode Khan, (2016). "Improving Security Of Keystroke Dynamics By Increasing The Distance Between Keys", *In proceeding of 3rd World Congress on Computer Applications and Information Systems* 2016, DOI: 08. WCCAIS.2016.1.10.

B. Bernardino, M. Vieira and D. Rasteiro, (2012). "Securing Data Warehouses from Web-Based Intrusions. *In Web Information Systems Engineering," LNCS 7651*, pages 681–688. Springer.

S. Bleha, C. Slivinsky, and B. Hussien, (1990). "Computer-Access Security Systems Using Keystroke Dynamics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, DOI: 10.1109/34.62613, vol. 12, pp. 1217-1222.

S. Cho, C. Han, D.H. Han, and H.L Kim, (2000). "Web-based keystroke dynamics identity verification using neural network, "*Journal of Organizational Computing and Electronic Commerce*, DOI: 10.1207/S15327744JOCE1004_07, vol. 10, No. 4, 295-307.

H. Davoudi, and E. Kabir, 2009. "A New Distance Measure for Fee Text Keystroke Authentication," *in the 14th International CSI Computer Conference*, pp. 570-575.

P. Dowland, and S.M Furnell, (2004). "A Long-term Trial of Keystroke Profiling using Digraph, Trigraph and Keyword Latencies," *Proceedings of IFIP/SEC 2004 - 19th International Conference on Information Security*, pp. 275-289.

R. Giot, B. Dorizzi, and C. Rosenberger, (2011). "Analysis of template update strategies for keystroke dynamics," *Proceedings of the IEEE Workshop on Computational Intelligence in Biometrics and Identity Management* (CIBIM '11); pp. 21–28.

I. Guyon and A. Elisseeff, (2003). "An introduction to variable and feature selection, *J. Machine Learning Research*," vol. 3, pp. 1157–1182, 944968.

M. Karnan, M and M. Akila, (2010) "Personal authentication based on keystroke dynamics using soft computing techniques," *Proceedings of the 2nd International Conference on Communication Software and Networks* (ICCSN '10); February 2010; pp. 334–338.

M. Karnan, M. Akila and A. Kalamani, (2009). "Feature subset selection in keystroke dynamics using ant colony optimization," *in Journal of Eng. and Tech. Research*, vol. 1(5), pp. 072- 080.

M. Karnan, M. Akila, and N. Krishnaraj, (2011). "Biometric personal authentication using keystroke dynamics," *a review. Applied Soft Computing Journal*. Vol. 11, No. 2, pp. 1565–1573.

K. Killourhy and R. Maxion, (2009). "Comparing anomaly-detection algorithms for keystroke dynamics," *IEEE/IFIP International Conference on Dependable Systems & Networks, DSN'09*, pp. 125–134, Jul. 2009.

F. Monrose, and A. Rubin, (1997). "Authentication via keystroke dynamics" *Proceedings of the 4th ACM Conference on Computer and Communications Security*; Zurich, Switzerland. pp. 48–56.

F. Monrose and A.D. Rubin, (2000). "Keystroke dynamics as a biometric for authentication, *Future Generation Computer Systems*, DOI: 10.1016/S0167-739X (99)00059-X, vol. 16 no. 4 pp. 351-359.

F. Monrose, M. Reiter, and S. Wetzel, (2002). "Password Hardening Based on Keystroke Dynamics", *International Journal of Information Heenaye-Mamode Security*. pp. 1-15.

A. K, Nag, D. Dasgupta, and K. Deb, (2014). "An Adaptive Approach for Active Multi-Factor Authentication", *9th Annual Symposium on Information Assurance* (ASIA'14), ALBANY, NY.

J.R Nisha, R.P. Anto Kumar, (2014). "User Authentication based on Keystroke Dynamics," *Journal of Engineering Research and Applications www.ijera.com* ISSN: 2248-9622, Vol. 4, Issue 3(Version 1), pp.345-349.

M.S Obaidat, and B. Sadoun, (1997). "Verification of computer users using keystroke dynamics", IEEE *Transactions on Systems, Man, and Cybernetics,—Part* B: DOI: 10.1145/75577.75582, vol. 27 no. 2, pp. 261-269.

L. O'Gorman, (2003). "Comparing Passwords, Tokens, and Biometrics for User Authentication". *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2019-2040.

S. Rajput, and P. Vijayawargiya, (2015). "Implementation of Keystroke Dynamics Application for Identifying Emotional State," *International Journal of Advanced Research in Computer Science and Software Engineering*.

B. Rao, (2005). "Continuous Keystroke Biometric System", Master's thesis, University of California.

A. Ross, and A.K Jain, (2004). "Multimodal Biometrics: An Overview," *Appeared in Proc. of 12th European Signal Processing Conference* (EUSIPCO), (Vienna, Austria), pp. 1221-1224.

J. Roth, X. Liu, A. Ross and D. Metaxas, (2014). "Investigating the Discriminative Power of Keystroke Sound," *in IEEE Transactions on Information Forensics and Security*, pp. 333-345.

K. Senathipathi and Batri, (2014) "An Analysis of Particle Swarm Optimization and Genetic Algorithm with Respect to Keystroke Dynamics," Green Computing Communication and Electrical Engineering (ICGCCEE), Coimbatore, pp. 1 – 11, Mar. 2014.

D. Shanmugapriya and G. A. Padmavathi, (2009). “Survey of Biometric keystroke Dynamics: *Approaches, Security and Challenges,” International Journal of Computer Science and Information Security*, Vol. 5, No. 1.

D. Shanmugapriya and G. A. Padmavathi, (2011) “An Efficient Feature Selection Technique for User Authentication using Keystroke Dynamics,” *IJCSNS International Journal of Computer Science and Network Security*, VOL.11 No.10, 191.

Z. Syed, and I. Syed, (2015). “Soft Biometrics for Keystroke Dynamics,” Universit´e de Caen Basse-Normandie, 2014. English. HAL Id: tel-01108638 https://hal.archives-ouvertes.fr/tel-01108638.

P.S Teh, S. Yue, and A.B.J Teoh, (2012). “Feature Fusion Approach on Keystroke Dynamics Efficiency Enhancement,” *International Journal of Cyber Security and Digital Forensics (IJCSDF)* 1(1), 20-31.

P.S Teh, A. Teoh, T.S Ong and H.F Neo, (2007). “Statistical Fusion Approach on Keystroke Dynamics,” Proceedings of the 3rd IEEE *International Conference on Signal Image Technologies and Internet Based Systems (SITIS ’07)*; pp. 918–923.

P. Teh, A.B.J Teoh, and S. Yue, (2013). “A Survey of Keystroke Dynamics Biometrics,” The Scientific World Journal Volume 2013 (2013), *Article ID 408280*, 24 pages.

K. Yongseong, W.N. Street and M. Filippo, (2000) “Feature selection in unsupervised learning via evolutionary search,” *ACM*, pp. 365–369, 347169.

Z. Zheng and L. Huan, (2007). “Spectral Feature Selection for Supervised and Unsupervised Learning,” *ACM*, pp. 1151–1157, 2007, 1273641.

7 NOTES ON CONTRIBUTORS

Purvashi Baynath has a degree in Electronics and Computer Science since 2011. She is actually pursuing her Ph.D. in the field of Keystroke Dynamics, Biometrics and Evolutionary Algorithms.

Professor Sunjiv Soyjaudah had a long and successful career at the University of Mauritius where he contributed significantly in academic research, namely in the fields of Communications Engineering and Computer Science and Engineering.

Maleika Heenaye-Mamode Khan is a senior lecturer at the University of Mauritius. She has a Ph.D. in Computer Science and Engineering. Her field of interest is biometrics, image processing and computer vision.