A Novel Approach for Keyboard Dynamics Authentication based on Fusion of Stochastic Classifiers

A. Rezaei and S. Mirzakochaki,

Iran University of Science and Technology Electronic Research Center

Summary
Computer security is one of most important issues around the world. Most computer systems are using passwords for their own authentication or verification mechanisms. A robust and efficacious approach for classification of 24 persons who their typing patterns were collected introduced. A linear discriminate classifier (LDC), quadratic discriminant classifier (QDC) and k-nearest neighbor (K-NN) are utilized to classify users keystroke patterns. After that a set of mentioned ensemble methods are adopted to reduce the error rate and increase the reliability of biometric authentication system. Promising results have been achieved. The best mean FAR, FRR and EER parameters are achieved for singular classifiers as 19.20%, 0.81% and 1.39% respectively. The state of the art performance results mean FAR, FRR and EER parameters are achieved for the ensemble classifiers as 0.00%, 0.00% and 1.15% respectively.

Key words:
Authentication, Imposter, keyboard dynamics, Linear discriminant classifier (LDC), quadratic discriminat classifer (QDC), k-nearest neighbor classifier (K-NN), Classifier combination, Discriminant functions.

1. Introduction

Nowadays computer systems are dominating traditional lifestyles. Banking, electronic purchasing, electronic elections, website subscription all and all require identification of a legitimate user. A traditional way and maybe the only choice nowadays is assigning a user name and a password to each user in order to authenticate them in the future. Each user can log into the system entering his/her username and password and access is granted if he/she enters the correct and match information.
Traditional password authentication scheme may be an effective way to authenticate legitimate users but it has several crucial drawbacks. Users easily forget their hard to guess username and password, write it down on a piece of paper or a computer file maybe near the terminal PC and share their own passwords with others due to some reasons. Passwords are prone to crack using brute force attacks or social engineering tricks [1,9], furthermore assigning superficial passwords will make the passwords prone to guess. A negligent user may use his/her birth date, passport number, social security number and so on as his/her password which is a blessing to imposters. Reassigning a new password to harden the security system also makes the user feel uncomfortable about memorizing a new hard to remind password. All of these drawbacks and vulnerability to imposters has made the password based authentication systems limited to non sensitive computer systems. Sensitive authentication systems may use an operator and some other attributes in order to minimize the intruding actions.
In order to harden and reinforce the authentication of computer systems and consequently the valuable users data, an alternative must be pondered. This alternative basically based on something the user have as rudimentary string based passwords, maybe some kind of authenticating devices like a electronic smart cards and finally some characteristics of the user which is known as a biometric [10].
A biometric is a physiological or behavioral feature of a living creature which makes it distinct from others by measuring them. In other words a biometric is a way of automated human recognition based on their physiological or behavioral characteristics [6]. Common physiological biometrics in use are hand fingerprint, palm geography detection, iris scan, palm scan, face detection and some more. In contrast as some instances of behavioral biometrics human gait, signature, keyboard stroke dynamics and voice detection can be named [2]. Biometric based authentication is very popular and prevalent all around the world.
Keyboard dynamics refers to as how a person types; In other words rhythm of typing is considered to discriminate between users. Recognition of users is done typically by capturing temporal or pressure profile of a person who types [11].
A brief preface has been covered in this section. Section 2 deals with a brief opening of keyboard dynamics authentication a. Section 3 reviews 3 works related to fusion of classification methods thoroughly. Section 4 introduces the features have been adopted to represent the discrimination of users patterns. Section 5 classifies the patterns and compares the performance results together and demonstrates which one of them performs better singularly. It also presents performance curves as a alternative and visual criterion of comparison. Section 6 discuss the combination methods and express the adopted methods briefly. Section 7 discusses around proposed singular and combined methods and compares the

Manuscript received August 5, 2012
Manuscript revised August 20, 2012

performance curves as well. Finally section 8 concludes the paper and discusses a brief review on the paper results.

2. Keyboard dynamics background

About 150 years ago in 1860's telegraph operators recognized each other using their key tapping patterns on telegraph keys. They easily could detect whether a specified operator has been substituted by another one. Traditional telegraph keys are extinct but modern keyboards are an array of the old ones. Decades after, an American company promulgated that keyboard stroke patterns of a typists have a relative consistency and therefore can be utilized as a behavioral biometric. Keyboard dynamics are proved to be a good technique to harden a password or PIN based authentication system. Consequently keyboard stroke patterns of a user can be affected by emotional states like anger, hunger, happiness, sorrow and so on which make the patterns deviate from normal template [2].

Keyboard dynamics authentication techniques generally scrutinize the user in two different styles: static and dynamic. In the static mode users are examined carefully just at the login time. After login process and granting access to the user there will not be any mechanisms to re-authenticate the same legitimate user.

Dynamic user authentication also called "free text authentication". In this mode, users typing keystroke patterns are continuously being examined. Naturally in this mode there will not be a predetermined password to type and authentication mechanism must examine the user using their typing patterns regardless of what is being typed. Logically in this category, verification of a user is not limited to login process and provides the system more robust mechanism to thwart intrudes. Most of researches has focused on static authentication rather that dynamic mode and there will be many gaps to fill to achieve a robust knowledge in free text field of research [4].

3. Literature review

There are a large number of researches in the area of keyboard or keystroke dynamics in order to harden the process of granting access to a claimed user. In some of them the approach is straightforward statistical methods, though in the rest the standard classification methods like linear discriminant functions (LDC), k-nearest neighbor and etc are utilized. In this section some scarce researches which adopted the concept of ensemble in keyboard dynamics authentication is reviewed thoroughly.

As pointed, the performed research in [12] can be mentioned that the researcher have made use of an artificial neural networks (ANN) classifier and a adaptive neuro-fuzzy interface system (ANFIS) in order to discriminate the users typing patterns. In the proposed system the ANN and ANFIS individually classifies the collected patterns from the users. The results from each of them are fused together using AND logic to make a decision. In other words the final result is positive just if both of the classifiers say yes to under test pattern.

In this works there is no results discussion between single classifiers and the ensemble and just the aggregate results are reported. The users patterns are collected using a numerical keyboard armed with pressure sensors and a data acquisition system (DAQ). The neural network in use is a multilayer perceptron (MLP) which is named as multilayer feed-forward neural network (MFN).

The utilized adaptive neuro-fuzzy interface system is compatible with Sugeno fuzzy rule-based logic. Selecting features from the collected dataset both temporal attributes like dwell time and flight time and exerted maximum pressure on each key are extracted from them. In this job 5 users under test have been dictated to enter a 6-character long password. Each of them are asked to login into the collecting system 200 times using their 6 character passwords. In order to improve system performance reporting the hold-out method has been used. Along with the 6 character passwords digraphs 5 temporal features and 6 pressure features have been extracted which forms an 11 feature problem. The resultant criterions for FAR equal to 0% from close set, 3.8% for open set and FRR equal to 0% are reported. According to the paper close set is related to legitimate user patterns and open set refers to as imposters.

There is another research by Shen et al. [13] which makes use of a relatively heuristic approach. The utilized classification method is nothing more than stochastic formulas. A set of relatively new definitions are adopted in this paper. In this research 50 users are asked to enter their favorite user name, password and a fixed admin-granted string 10 times. From the evidence it can be inferred that this job is in category of password hardening and static user recognition techniques. In this job two methods of scoring are used to discriminate between user patterns of typing. In the first method similarity scoring is utilized using a Gaussian function between under test pattern and the template pattern.

In the second method direction similarity measure (DSM) is utilized in order to compare under test patterns and the template. Concerning a threshold value the mentioned methods can be formed to stochastic classifiers. Finally these two decisions are combined together and final decision is made. The combination rule would be weighted summation. The ensemble results demonstrate better improvements in contrast to single methods. It must be mentioned that Gaussian scoring would have higher weight or value that the direction similarity measure method. The mentioned Gaussian function in this job is offered in Equation. 1.

$$f(D_i, \sigma_i, \mu_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\left(\frac{(D_i - \mu_i)^2}{2\sigma_i^2}\right)} \quad (1)$$

Using Eq. 1 score of feature i is evaluated and the parameters σ_i and μ_i are calculated in accordance with each user typing pattern. Gaussian final score is evaluated from mean of all scores from each individual feature score. Equation .2 demonstrates the relationship this relationship.

$$\text{Score}_{(GD,Total)} = \frac{\sum_{i=1}^{k} \text{Score}_{(GD,D_i)}}{k} \quad (2)$$

In the second proposed heuristic method, scoring is performed using the mean vector for each collected features and comparison of each digraph gradient with corresponding value in the mean vector. If the digraph gradient in under test vector and the mean vector is a positive value, a counter names m is incremented else it is leaved unchanged. Finally this score is calculated using Equation. 3. In this equation n is the number of all characters in the typed string.

$$\text{Score}_{(DSM,Total)} = \frac{m}{n-1} \quad (3)$$

The ensemble classifier is formed using a weighted summation of scores of the former and the latter methods. In the proposed weighted summation, w is ranged between 0 and 1 and would be chosen by the administrator. Assigning a value for threshold makes the ensemble scoring weights as a bivalent classifier. Evaluating the ensemble EER of 11.68% is achieved roughly from Gaussian measure, 19.74% from direction similarity measurement and 9.96% for the ensemble with the most proper value for the weight w.

Another notable research by Hocquet et al. [14] has used the idea of combinng three classifiers in order to reach to lower classification errors. In this research keyboard dynamics of 15 people during 6 month has been collected.

In the first proposed method 17 users are asked to type a paragraph including roughly 1000 keystrokes into collecting software. Mean and standard deviation of the collected features are evaluated and has been used for comparing to under test features. In the case at least 60% of match is detected the patterns are supposed to be from a legitimate user. The criterion used for matching the under test patterns and the template is defined as relation $|x_i - \mu_i| \leq \frac{1}{2}\sigma_i$. FRR and FAR parameters are evaluated as 5% and 5.5% respectively. In contrast to some other researches in this field it is suggested to implement a scoring measure like the former job. In the second proposed method the some categories are related to the temporal ranges. These ranges have been divided to very short, short, moderate, long and very long. In the second method the disorders between vectors are used. Each collected feature is ranked descendingly in order to classify the patterns. In addition the mean value for each user is stored. Assigning a score to each pattern is performed using Euclidean distance between under test and the template or Spearman's coefficients calculation. At last trinary methods are combined together to make an ensemble using minimum, maximum, median, product, majority vote, unique voice and summation methods. The best ensemble results are reported as 1.17% for FAR, 0.54% for FRR and 1.75% for ERR in summation rule. Losing simplicity after ensemble the singular classifiers a better error results are achieved.

4. Feature extraction:

QWERTY keyboards are in use in many personal computers nowadays. They are ubiquitous as most of computer systems make use of ordinary QUERTY keyboards. There is no possibility to gain the force each user exerts to do the typing. Some researcher has performed thorough investigation using force sensitive keyboard successfully like [5].

Keyboard dynamics features can be divided into to integral characteristics: global and temporal. Global features is related to general and common habits of a specified user like typo frequency, utilizing right or left control keys like Alt, Ctrl and Shift keys, utilizing number pad for typing numbers or the alternative keys, and aggregate typing speed. The temporal features reflect the habitual typing style of a specified user which might be timing of keystrokes or key-press forces exerted during typing [7].

Using ubiquitous keyboards, the only measure we can extract from a user pattern would be timing of strokes. In other words a sniffer application must extract each performed action on the keyboard and log it. For instance Fig. 1 demonstrates the timing sequence of typed word "KEY".

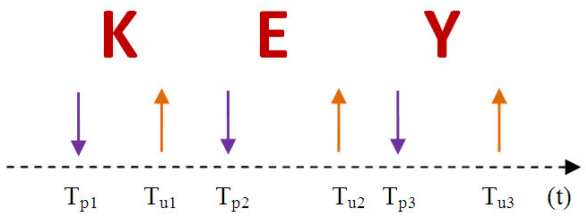

Fig.1: Temporal timing features of a sample typed string "KEY"

Versatile timing features can be extracted from Fig.1 to represent timing habits of a defined user. Basically hold time (dwell time) can be defined the time interval between pressing a specified key and releasing the same key for instance for pressed key "E" T_{u2}-T_{p2} is defined as hold interval. Accordingly inter-key time (flight time) can be defined the time interval between pressing a key and pressing the next key or releasing a key and releasing the next one. For typed sequence of "EY" the time interval T_{p3}-T_{p2} can be regarded [8].

A timer enable application can store all the time transactions and log it on the database. For decreasing the deviation of typing patterns of users maybe several same typing strings are asked from users. Increasing samples of features from users helps classification training process

more robust and consequently less testing errors would occur.

5. Singular classification of collected patterns:

In this section a brief review of mathematical aspects of 3 proposed classifiers are given. Linear discriminant classifier (LDC), quadratic discriminant classifier (QDC) and k-nearest neighbors classifier (K-NN) are introduced briefly. These classifiers are used furthermore to make an ensemble.

5.1. Linear discriminant classifier (LDC):

The name linear classifier is derived from type of discriminant functions the classifier makes use. In this kind of classifier any set of linear functions like $g_i: R^n \rightarrow R, i = 1, \dots, c$ can be used as a linear classifier. These kinds of discriminant functions can be described as Equation. 1.

$$g_i(x) = w_{i0} + w_i^T x, \quad x, w_i \in R^n, w_{i0} \in R \tag{1}$$

Linear classifiers are of simplest forms of classifiers ever used and have a powerful background in the literature. Linear discriminant classifiers are straightforward to evaluate and have a robust nature to discriminate patterns even when classes do not have normal distributions. Rearranging the maximum membership rule as a rule to decide which class label to choose results in Equation. 2.

$$D(x) = \omega_{i*} \in \Omega \leftrightarrow P(\omega_{i*}|x) \tag{2}$$

Given the prior probabilities and class conditional probability density functions (PDF) also knowing the Bayesian rule formula from conditional probability as stated in Equation. 3. it is easy to evaluate the posterior probabilities where c is number of classes.

$$P(\omega_i|\boldsymbol{x}) = \frac{P(\omega_i)p(\boldsymbol{x}|\omega_i)}{p(\boldsymbol{x})} = \frac{P(\omega_i)p(\boldsymbol{x}|\omega_i)}{\sum_{j=1}^{c} P(\omega_j)p(\boldsymbol{x}|\omega_j)} \tag{3}$$

Due to equality of the denominator of Equation. 3 for all of class labels ω_i and assuming $p(x)$ is nonnegative value so ranking order of discriminant functions $g_i(x)$ will not alter. A set of discriminant functions which lead to the same result of classification as Equation. 3 suggests can be expressed using Equation. 4.

$$g_i(x) = P(\omega_i|x) \propto P(\omega_i)p(x|\omega_i), \quad i = 1, \dots, c \tag{4}$$

In addition a logarithmic version of Equation. 5 can be in use in the corresponding literature as Equation. 5 suggests.

$$g_i(x) = \log [P(\omega_i)p(x|\omega_i)], \quad i = 1, \dots, c \tag{5}$$

The class with highest posterior probability is the most accepted and logical class to choose in the face of pattern **x.** Assuming all classes are Gaussian distributed with means μ_i and covariance matrices Σ_i and also assuming $p(x|\omega_i)$ can be evaluated with a statistical distribution consisting of mean matrix of μ_i and covariance matrix of Σ_i for all classes or i=1,...,c Equation. 5 reshapes and will be converted to Equation. 6.

$$g_i(x) = log[P(\omega_i)] + \tag{6}$$

$$log\left\{\frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma_i|}} e^{\left[-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right]}\right\}$$

Utilizing mathematic characteristics of logarithm function, Equation. 6 can be simplified as Equation. 7 suggests.

$$g_i(x) = log[P(\omega_i)] - \frac{n}{2}log(2\pi) - \frac{1}{2}log(|\Sigma_i|) \tag{7}$$
$$-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)$$

Assuming that class covariance matrices Σ_i are the same and equals to Σ, class conditional probability density function is distributed normally with mean μ_i and covariance Σ_i and discarding the terms which does not relate to class ω_i leads in simplest discriminant function which still does not change the ranking order of prior discriminant functions $g_i(x)$. Finally the new generated discriminant functions $g_i(x)$ can be introduced as Equation. 8.

$$g_i(x) = log[P(\omega_i)] - \frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \mu_i^T \Sigma^{-1} x = \tag{8}$$
$$w_{i0} + w_i^T x$$

Evidently it is inferred from the equation. 8 that term without pattern x multiplied can be regarded as w_{i0} and the term has multiplying pattern x at the end can be in the role of term w_i^T. In Equation. 8 $w_{i0} \in R$ is a real number and $w_i \in R^n$ is a row vector that plays role as coefficients and describes the linear discriminant classifier discriminant functions. Being sanguine nor the classes of a dataset are distributed normally nor the real values of μ_i and Σ_i are known to us. In reality an approximation of mean and covariance matrices are utilized which guarantees the obtained classifier will not have minimum error [3].

5.2. k-nearest neighbors classifier (k-NN):

K-nearest neighbors classifier is one of the most straightforward and prevalent classifiers ever known in pattern recognition. In designing non-parametric classifiers probability density function $p(x|\omega_i)$ is approximated near under test pattern in R^n. If we name this region in R^n as R, approximate probability $\hat{p}$ can be formed by Equation 9.

$$\hat{p} = P(x \in R) = \int_R p(w)dw \tag{9}$$

$\hat{p}$ can be described roughly by dividing number of patterns available in R by the whole number of patterns N as Equation. 10 demonstrates this fraction.

$$\hat{p} \approx \frac{k}{N} \tag{10}$$

Assuming Pattern x is located in R, for small enough regions of R p(w) or probability function in Equation. 10 can be supposed constant and Equation.10 reshapes to Equation. 11. In new Equation V_R is related to volume of $R \in R^n$.

$$\hat{p} \approx \hat{p}(x) \int_R dw = \hat{p}(x).V_R \tag{11}$$

Combining Equation 10 and 11 yields $\hat{p}(x)$ as Equation. 12 demonstrates it.

$$\hat{p}(x) \approx \frac{\mathrm{k}}{\mathrm{N}.V_R} \tag{12}$$

In the case $N \to \infty$ and $V_R \to 0$ the approximated probability function $\hat{p}(x)$ tends to get to real $p(x)$. Assuming k_i patterns with class label ω_i is available in region R, the probability density function for each of class labels ω_i in region R can be approximated using Equation. 13. $\mathrm{N_i}$ in this equation represents the number of patterns with class label ω_i in region R.

$$p(x|\omega_i) \approx \frac{\mathrm{k_i}}{\mathrm{N_i}.V_R} \tag{13}$$

Putting evaluated probabilities in Equations 12 and 13 and calculating prior probability $p(\omega_i)$ on assumption $\mathrm{N_i}$ patterns are put in class ω_i and N represents the number of all patterns, posterior probability based on Bayes rule for class label ω_i would be evaluated in Equation. 14.

$$p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{p(x)} \approx \frac{\frac{\mathrm{k_i}}{\mathrm{N_i}.V_R}\frac{\mathrm{N_i}}{\mathrm{N}}}{\frac{\mathrm{k}}{\mathrm{N}.V_R}} \tag{14}$$

Finally simplifying Equation. 14 yields Equation. 15.

$$p(\omega_i|x) = \approx \frac{\mathrm{k_i}}{\mathrm{k}} \tag{15}$$

In other words posterior probability for class label ω_i can be expressed as ratio of pattern with label ω_i in region R to whole patterns in region R [3].

5.3. Quadratic discriminant classifier (QDC):

As stated in section 5.1 linear discriminant classifiers are composed of a linear hyperplane in $\mathrm{R^n}$ which discriminates the patterns. Equation. 16 states linear discriminant classifier discrimination function.

$$\mathrm{g(x)} = \mathrm{w_0} + \sum_{\mathrm{i=1}}^{\mathrm{d}} \mathrm{w_i x_i} \tag{16}$$

By adding an extra term including product of different features and proper weights, a new nonlinear classifier discriminant function is resulted. This nonlinear classifier is named quadratic discriminant classifier (QDC). In Equation. 16 and Equation. 17 d represent number of features in pattern x. As Equation. 17 states no more linear hyperplane exists to discriminate the patterns in $\mathrm{R^n}$ and the problem takes a second order from [15].

$$\mathrm{g(x)} = \mathrm{w_0} + \sum_{\mathrm{i=1}}^{\mathrm{d}} \mathrm{w_i x_i} + \sum_{\mathrm{i=1}}^{\mathrm{d}} \sum_{\mathrm{j=1}}^{\mathrm{d}} \mathrm{w_{ij} x_i\, x_j} \tag{17}$$

6. Classifiers combination:

Combining classifiers is a powerful method for increasing classification performance in pattern recognition complicated problems. In many applications it is empirically shown that using ensemble of classifiers is worth using one in the cost of loss of simplicity and time.

In general there are two methods for classifiers ensemble: classifier selection and classifier fusion. In selection it is assumed that each classifier performs well in some features space of under test patterns. In this method the final result is extracted from a single classifier or a selection of them. In fusion method it is assumed that all classifiers are trained in features space and they acts as complementary rather competition [16].

Fusion of a set of stochastic classifiers itself can be performed in two ways: trainable and non-trainable. In trainable way of fusion the results adaption is used to gain maximum performance. Although in non-trainable classifiers the parameters are constants and are not a function of classification training results. In other words classifiers are united using a single criterion after training [16].

Assuming a set of classifiers $\mathrm{D} = \{\mathrm{D_i}|\mathrm{i} \in 1, \dots, \mathrm{L}\}$ and also a set of class labels $\omega = \{\omega_\mathrm{i}|\mathrm{i} \in 1, \dots, \mathrm{c}\}$ as the results set for the mentioned classifiers, $\mathrm{D_i}$ returns decision profile matrix (DP matrix) $\{\mathrm{d_{ij}}|\, 0 \le \mathrm{d_{ij}} \le 1, \mathrm{j} \in 1, \dots, \mathrm{c}\}$ in the case of under test pattern x and for each class label. In other words $\mathrm{d_{ij}}$ is degree of support of classifer $\mathrm{D_i}$ from class label ω_j. According to maximum membership rule class label ω_k is introduced as output of classifier $\mathrm{D_i}$ in the case for all $\mathrm{j} \in 1, \dots, \mathrm{c}$ inequation $\mathrm{d_{ik}} \ge \mathrm{d_{ij}}$ holds.

In this matrix support degree vector from decision which is made of class ω_j is dedicated in row i. Easily deduced that column j in DP matrix is dedicated to support degree of class label ω_j from all set of classifiers.

Defining combiner function $\mathcal{F}_\mathrm{j}[\mathrm{d_{ij}}]$ $\mathrm{i} \in 1, \dots, \mathrm{L}$ on column j of DP matrix the ensemble support degree is achieved for the respective function. There will be infinite choices for the combiner classifiers but some kinds of them are the most prevalent. These functions are listed in Equation. 18 to Equation. 22.

$$\mu_{j_{min}}(x) = \min_i(\mathrm{d_{ij}(x)}) \tag{18}$$

$$\mu_{j_{max}}(x) = \max_i(\mathrm{d_{ij}(x)}) \tag{19}$$

$$\mu_{j_{mean}}(x) = \frac{1}{L}\sum_{i=1}^{L} \mathrm{d_{ij}(x)} \tag{20}$$

$$\mu_{j_{median}}(x) = \underset{i}{median}(\mathrm{d_{ij}(x)}) \tag{21}$$

$$\mu_{j_{product}}(x) = \prod_{i=1}^{L} \mathrm{d_{ij}(x)} \tag{22}$$

7. Experimental results:

Dataset collection is based on extracting selected proper features from 24 legitimate users patterns during 2 sessions of data collection. Overall 100 patterns for each user are collected and stored in the database. Training of each single classifier is based on 70% of collected data; In other words 70 patterns of each user is utilized for training and the rest 30 patterns are utilized for testing the performance of each classifier or the ensemble of them.

The main features of typing pattern for a fixed (static) password were investigated in section 4. In this investigation no discriminative difference or weight between temporal features are considered in classification process. The classification process has been performed on 2400 collected patterns.

Each of proposed classifiers is applied and the performance criterions are achieved. Relatively good and acceptable results have been achieved. As pointed before Linear discriminant classifier math assumes that normal distribution is around each class though Gaussian class dependent probability density function can be assumed. Regardless of examining the Gaussian distribution among features linear discriminant classifier has displayed acceptable results. Table. 1 demonstrates the performance results thoghrouly.

It must be noted here that as several parameters like collection methods, database patterns per each class, number of features, error evaluation method, and maybe multiple error criterion and lack of a coherent workbench has make the comparison of different researches impossible and illogical. But many researchers have done so. So other researches results are overlooked in the table.

7.1. Single Classifiers performance results:

In this section error evaluation of each introduced single classifier has been made. All mentioned values are considered in percent except area under curve (AUC). The performance results express promising values for FRR but still have relatively tangible values for FAR. FAR criterion has a more important impact on biometric system rather than FRR. That is because rejecting a legitimate user from accessing the system just has a mental impact on users and does not have any loss of valuable data for imposters. Error of each classifier has been evaluated using π-method or cross-validation method to ensure that the results are trustable. As EER values demonstrate overall performance of a classifier it is deduced from Table. 1 that K-NN has lower performance rather than LDC and QDC but it still has lower area under curve parameter rather than the rest. K-NN also expresses higher FAR parameters which makes it a improper choice singularly for a biometric authentication system.

Table. 1: Performance comparison among single classifiers LDC, QDC and K-NN

Classifier	FAR (Mean)	FAR (Min)	FAR (Max)	FRR (Mean)	FRR (Min)	FRR (Max)
LDC	19.20	6.67	36.67	0.81	0.14	1.74
QDC	27.20	10.00	46.67	1.39	0.14	3.91
KNN	39.23	23.23	56.67	1.55	0.58	3.48
Classifier	**ERR (Mean)**	**ERR (Min)**	**ERR (Max)**	**Error (π-method)**		**AUC**
LDC	2.98	2.56	3.77	11.43		8.38
QDC	1.39	1.00	8.62	11.46		3.58
K-NN	5.87	6.02	12.26	11.24		2.76

Visually comparing the performance results as Fig. 2 suggests it can be inferred that LDC has the most proper ROC rather than the rest. It must be noted that ROC curves are illustrated in logarithmic scale to mention the results more distinct rather than linear scale.

In addition to ROC curves which illustrate the performance of classifiers, respective learning curves for singular classifiers are drawn in Fig. 3. These graphs shows a single classifier how much can learn according to number of learning pattern introduced to the training stage of them. As it can be deduced from Fig. 3 each of triple classifiers relatively learns like each other and no tangible discrimination can be visible.

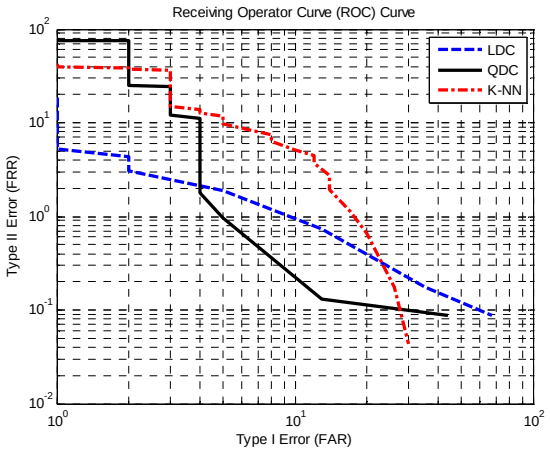

Fig. 2: ROC curves for single classifiers on collected keyboard dynamics patterns

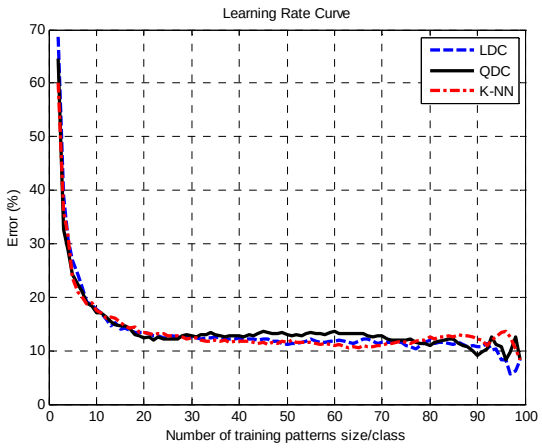

Fig. 3: Learning rate curves for single classifiers on collected keyboard dynamics patterns

7.2. Ensemble Classification performance results:

Relatively acceptable results extracted from single classifiers as stated in section 7.1. In this section a

thorough examination has been performed on combined classifiers. Table.2 illustrates thorough investigation on performance parameters of the ensemble of triple classifiers. As Table. 2 denotes relative error criterions have reduced to a more acceptable range. In the case of FRR all ensemble methods has shown betterments on this parameter. In the case of FAR it is the same. Rigorously it can be deduced from the results that combination methods has made the authentication system more robust than its previous performance. In the case of EER the ensemble results are relatively in the range of singular classifiers except minimum method of combination which has

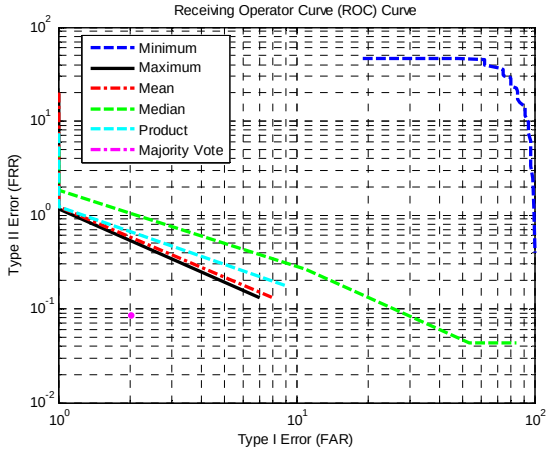

Fig. 4: ROCs for combined classifiers on collected keyboard dynamics patterns

showed an anomaly in contrast to other combination method like maximum and mean.

The discussion is also promising for the mean error evaluation. The overall mean errors for ensemble classifiers have been abated tangibly in comparison to singular classifiers as Fig. 10 states. In the case of cross-validation the error results are relatively in the same range and does not displays betterment as Fig. 12 states. Another important parameter which in almost all of research reports has been ignored reporting it is the time each classifier takes to be trained and evaluated. In this research a rough elapsed time for each classifier and ensemble methods has been introduced in Fig. 9. Timing parameters are very crucial in real authentication systems because it dictates the time a classifier takes when a new user appends to current dataset.

Table. 2: Performance comparison among different combination methods

	FAR (Mean)	FAR (Min)	FAR (Max)	FRR (Mean)	FRR (Min)	FRR (Max)
min	0.00	0.00	0.00	0.00	0.00	0.00
max	2.77	0.00	10.00	0.21	0.00	0.72
mean	4.13	0.00	13.33	0.22	0.00	0.58
median	10.73	0.00	23.33	0.53	0.14	1.01
product	7.23	0.00	16.67	0.30	0.00	0.72
Majority	1.77	0.00	6.67	0.12	0.00	0.29

	ERR (Mean)	ERR (Min)	ERR (Max)	Error (π-method)	AUC
min	46.52	46.52	46.52	11.51	0.04
max	1.15	1.15	1.15	11.49	3.91
mean	1.19	1.19	1.19	11.57	3.91
median	1.46	1.46	1.46	11.41	5.00
product	1.19	1.19	1.19	11.56	4.28
Majority	1.84	1.84	1.84	11.57	0.01
min	46.52	46.52	46.52	11.51	0.04

Visually comparing the performance results as Fig. 4 suggests it can be inferred that LDC has the most proper ROC rather than individual ones. It must be noted that ROCs are illustrated in logarithmic scale to mention the results more distinct rather than linear scale.

In addition to ROC curves which illustrate the performance of a classifier, respective learning curves for singular classifiers are drawn in Fig. 5. These graphs shows a single classifier how much can learn according to number of learning pattern introduced to the training stage of them. As it can be deduced from Fig. 5 the combined classifiers relatively learn like each other and no tangible discrimination can be visible.

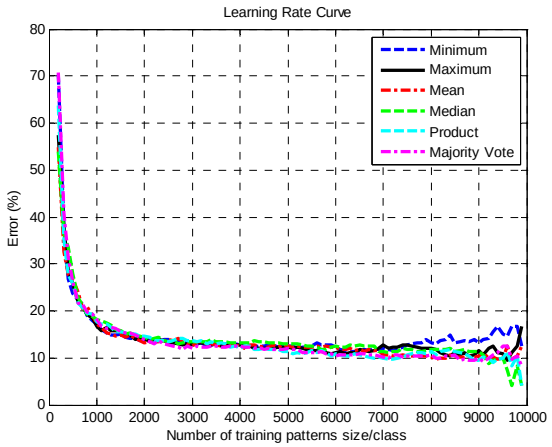

Fig. 5: Learning rate curve for combined classifiers on collected keyboard dynamics patterns

In addition to these illustrations the FAR, FRR and ERR parameter are denoted in Fig. 6, 7 and 8 respectively. It is to say minimum ensemble method has made the mean FAR and mean FRR zero which is a great improvement which never reported before.

It must be noted that parameters which mentioned are evaluated in mean evaluated in 100 round training process which make the mean values more reliable than a single round. Just the cross-validation error or π-method due to its nature of multiple testing has been evaluated once in 100 round. In addition area under curve graphs have shown a slight error for minimum and majority-vote combination methods as Fig. 11 denotes it.

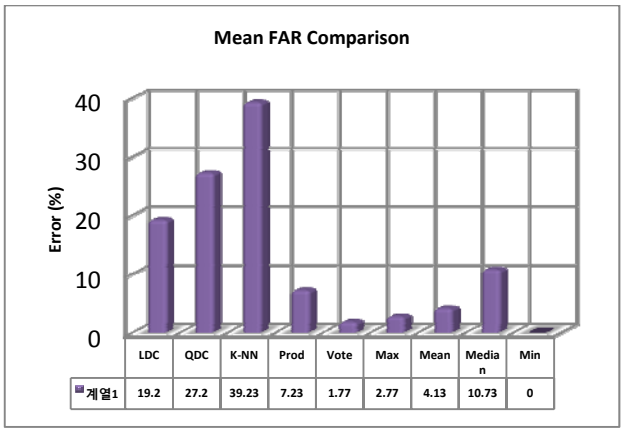

Fig. 6: FAR comparison among singular classifiers and ensemble ones

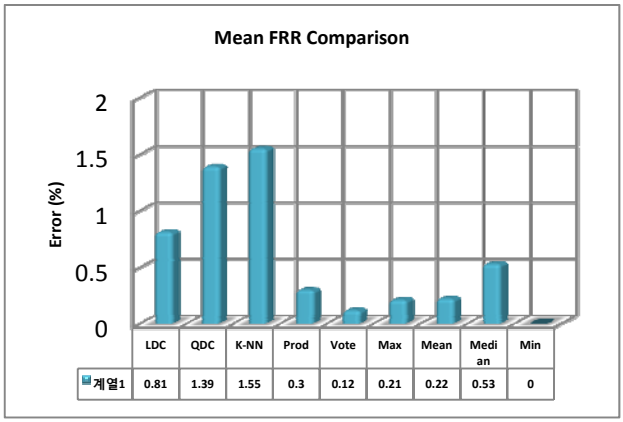

Fig. 7: FRR comparison among singular classifiers and ensemble ones

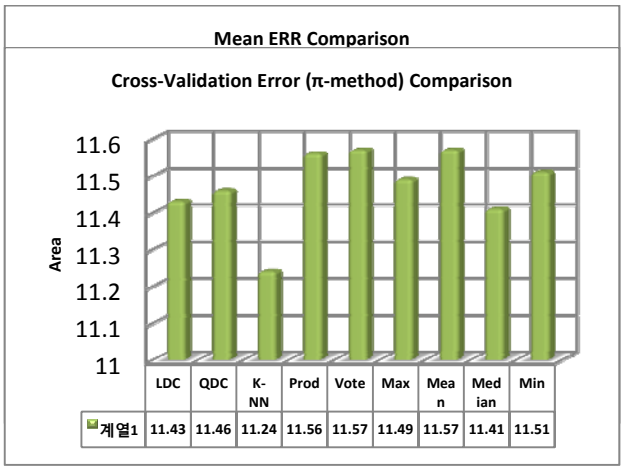

Fig. 12: Cross-Validation Error (π-method) Comparison among singular classifiers and ensemble ones

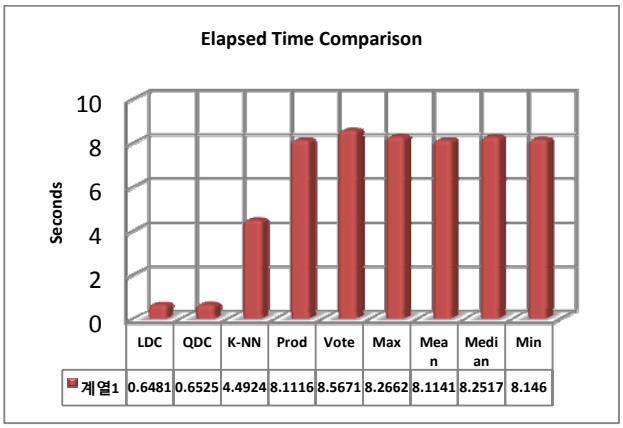

Fig. 9: Elapsed time for training and evaluating the patterns comparison among singular classifiers and ensemble ones

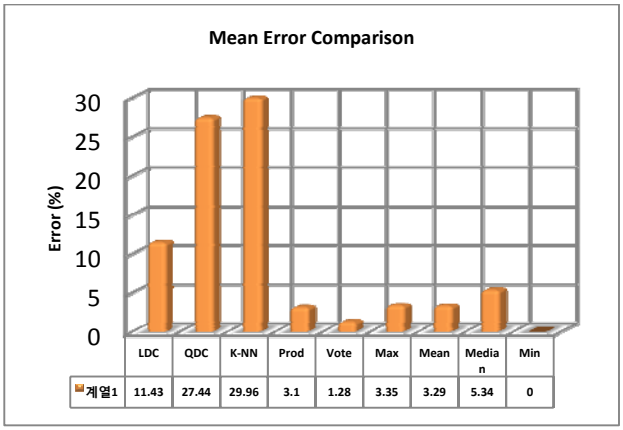

Fig. 10: Mean error comparison among singular classifiers and ensemble ones

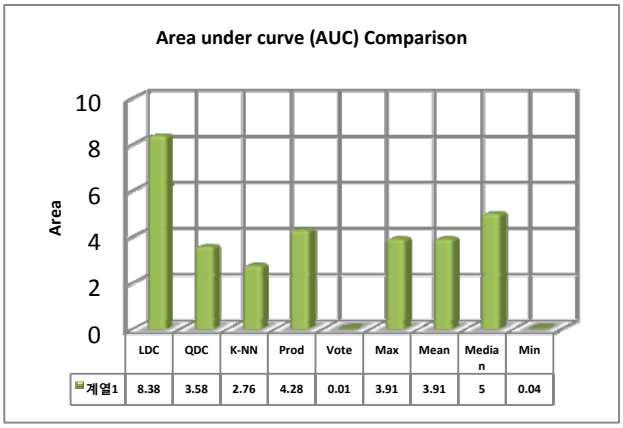

Fig. 11: Area under curve (AUC) comparison among singular classifiers and ensemble ones

8. Conclusion

A robust and efficient biometric authentication system is proposed using users keyboard dynamics patterns. The patterns are collected on 24 users each entering 100 patterns on a fixed 10 character passwords. Temporal features are extracted using proper definition. Overall 3 classifiers namely linear discriminant classifier (LDC), quadratic discriminant classifier (QDC) and K-nearest neighbor are trained individually on the collected pattern database. The acceptable results are achieved on the dataset. The best mean FAR, FRR and ERR parameter is achieved as 19.20%, 0.81% and 1.39% for singular classifiers. Combining individual classifiers based on minimum, maximum, mean, median, product and majority vote functions are implemented successfully and promising results have been achieved. Delving into the results minimum function has granted 0.00% error for both mean FAR and mean FRR. It also showed very small area under curve (AUC) near 0.04 which is negligible in contrast to other combiner functions. In the case of timing considerations K-NN has taken relatively large values in comparison with the 2 other classifiers. The ensemble best mean FAR, FRR and ERR parameter is achieved as 0.00%, 0.00% for minimum function and 1.15% for maximum function.

The keystroke dynamics authentication based on single classifications still has a long way to get matured due to relatively large values of FAR. But combining just 3 classifiers has demonstrated that error criterions for 24 persons can be reduced as little as zero. It means combing classifiers makes the biometric system so robust that low level of error parameters have been achievable.

References

[1] Jagadeesan, H.; Hsiao, M.S.; , "A novel approach to design of user re-authentication systems," Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on , vol., no., pp.1-6, 28-30 Sept. 2009

[2] Pin Shen Teh; Teoh, A.; Thian Song Ong; Han Foon Neo;, "Statistical Fusion Approach on Keystroke Dynamics," Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on , vol., no., pp.918-923, 16-18 Dec. 2007

[3] Ludmila I. Kuncheva, "Combining Pattern Classifiers: Methods and Algorithms", John Wiley & Sons Publications, July 2004, pp. 45-57.

[4] Davoudi, H.; Kabir, E.; , "A new distance measure for free text keystroke authentication," Computer Conference, 2009. CSICC 2009. 14th International CSI , vol., no., pp.570-575, 20-21 Oct. 2009

[5] De Ru, W.G.; Eloff, J.H.P.; , "Enhanced password authentication through fuzzy logic," IEEE Expert , vol.12, no.6, pp.38-45, Nov/Dec 1997. [6] B. Miller, "Vital signs of identity", IEEE Spectrum, 1994, pp. 22-30.

[6] Shimshon, T.; Moskovitch, R.; Rokach, L.; Elovici, Y.; , "Clustering di-graphs for continuously verifying users according to their typing patterns," Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of , vol., no., pp.000445-000449, 17-20 Nov. 2010.

[7] Sim, T.; Janakiraman, R.; , "Are Digraphs Good for Free-Text Keystroke Dynamics?," Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on , vol., no., pp.1-6, 17-22 June 2007, pp. 2.

[8] Budi Arief and Denis Besnard, "Technical and Human Issues in Computer-Based Systems Security", EICAR 2005 Conference, 2003, pp. 5.

[9] Hu, J.; Gingrich, D.; Sentosa, A.; , "A k-Nearest Neighbor Approach for User Authentication through Biometric Keystroke Dynamics," Communications, 2008. ICC '08. IEEE International Conference on , vol., no., pp.1556-1560, 19-23 May 2008.

[10] D. Shanmugapriya, G. Padmavathi, "A Survey of Biometric keystroke Dynamics: Approaches, Security and Challenges", (IJCSIS) International Journal of Computer Science and Information Security,Vol. 5, No. 1, 2009. pp. 1.

[11] Carlos, J.; Garcia, F.; Mendez, J.J.S.; , "A Comparison of ANFIS, ANN and DBR systems on volatile Time Series Identification," Fuzzy Information Processing Society, 2007. NAFIPS '07. Annual Meeting of the North American , vol., no., pp.319-324, 24-27 June 2007.

[12] Pin Shen Teh; Teoh, A.; Thian Song Ong; Han Foon Neo; , "Statistical Fusion Approach on Keystroke Dynamics," Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on , vol., no., pp.918-923, 16-18 Dec. 2007.

[13] Hocquet, S.; Ramel, J.-Y.; Cardot, H.; , "Fusion of methods for keystroke dynamic authentication," Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on , vol., no., pp. 224- 229, 17-18 Oct. 2005.

[14] Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern classification", Wiley, 2001, pp. 218.

[15] Rachid Benmokhtar, Benoit Huet, "Classifier Fusion: Combination Methods For Semantic Indexing in Video Content", In Proceedings of ICANN (2), 2006, pp. 65-66.