UDC 004.93

KEYSTROKE DYNAMICS RECOGNITION USING NINE-VARIATE PREDICTION ELLIPSOID FOR NORMALIZED DATA

Prykhodko S. B. – Dr. Sc., Professor, Head of the Department of Software for Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

Trukhov A. S. – Post-graduate student of the Department of Software for Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

ABSTRACT

Context. Keystroke dynamics recognition is a crucial element in enhancing security, enabling personalized user authentication, and supporting various identity verification systems. This study investigates the influence of data distribution on the performance of one-class classification models in keystroke dynamics, focusing on the application of a nine-variate prediction ellipsoid. The object of research is the keystroke dynamics recognition process. The subject of the research is a mathematical model for keystroke dynamics recognition. Unlike typical approaches assuming a multivariate normal distribution of data, real-world keystroke datasets often exhibit non-Gaussian distributions, complicating model accuracy and robustness. To address this, the dataset underwent normalization using the multivariate Box-Cox transformation, allowing the construction of a more precise decision boundary based on the prediction ellipsoid for normalized data.

The objective of the work is to increase the probability of keystroke dynamics recognition by constructing a nine-variate prediction ellipsoid for normalized data using the Box-Cox transformation.

Method. This research involves constructing a nine-variate prediction ellipsoid for data normalized using the Box-Cox transformation to improve keystroke dynamics recognition. The squared Mahalanobis distance is applied to identify and remove outliers, while the Mardia test assesses deviations from normality in the multivariate distribution. Estimates for parameters of multivariate Box-Cox transformation are derived using the maximum likelihood method.

Results. The results demonstrate significant performance improvements after normalization, reaching higher accuracy and robustness compared to models built for non-normalized data. The application of the nine-variate Box-Cox transformation successfully accounted for feature correlations, enabling the prediction ellipsoid to better capture underlying data patterns.

Conclusions. For keystroke dynamics recognition, a mathematical model in the form of the nine-variate prediction ellipsoid for data normalized using the multivariate Box-Cox transformation has been developed, which enhances the probability of recognition compared to models constructed for non-normalized data. However, challenges remain in determining the optimal normalization technique and selecting the significance level for constructing the prediction ellipsoid. These findings underscore the importance of careful feature selection and advanced data normalization techniques for further research in keystroke dynamics recognition.

KEYWORDS: keystroke dynamics recognition, multivariate Box-Cox transformation, prediction ellipsoid, normalizing transformation.

ABBREVIATIONS

BCT is the Box-Cox transformation;
SMD is the squared Mahalanobis distance;
TP is true positives;
FP is false positives;
TN is true negatives;
FN is false negatives;
NGD is a prediction ellipsoid for non-Gaussian data;
ND is a prediction ellipsoid for normalized data.

NOMENCLATURE

p is a number of variables;
m is a number of degrees of freedom;
N is a number of data points;
$\mathbf{S_X}$ is a sample covariance matrix for initial data;
$\mathbf{S_Z}$ is a sample covariance matrix for normalized data;
$\mathbf{X}$ is a non-Gaussian random vector;
$\overline{\mathbf{X}}$ is a vector of sample means of the X_j variables;
X_j is a j-th non-Gaussian variable;
$\mathbf{Z}$ is a Gaussian random vector;
$\overline{\mathbf{Z}}$ is a vector of sample means of the Z_j variables;
Z_j is a j-th Gaussian variable that is obtained by transforming the variable;
α is a significance level;
β_1 is a multivariate skewness;
β_2 is a multivariate kurtosis;
$\chi^2_{m,\alpha}$ is the chi-square distribution quantile with m degrees of freedom and significance level α;
$\boldsymbol{\psi}$ is a vector of multivariate normalizing transformation.

INTRODUCTION

In recent years, keystroke dynamics, also known as keystroke biometrics or typing biometrics, has emerged as a viable method for biometric authentication. It leverages the unique patterns and rhythms individuals exhibit while typing on a keyboard, capturing characteristics such as keystroke duration and inter-key intervals [1]. These features enable the creation of a distinctive typing profile for each user.

Unlike traditional biometric methods such as fingerprints or facial recognition, keystroke dynamics offers a non-intrusive and continuous form of user authentication

DOI 10.15588/1607-3274-2025-1-9

[2]. This makes it particularly attractive for applications in online banking, secure login systems, and access control. A low probability of authenticating individuals can have negative consequences in terms of security and personalization. Therefore, there is a need to develop and improve keystroke dynamics recognition methods.

A significant limitation of existing methods in keystroke dynamics recognition is their sensitivity to non-normal data distribution. Many statistical and machine learning techniques assume multivariate normality in the underlying data, which is often not the case in real-world keystroke dynamics datasets [4]. Non-normal distributions can adversely affect the performance of these techniques, leading to suboptimal recognition results. Therefore, it is necessary to enhance mathematical models to accommodate deviations from the multivariate normal distribution of data.

The object of study is the process of keystroke dynamics recognition.

The keystroke dynamics recognition process includes several important steps to ensure accurate and reliable user authentication. First, a suitable dataset containing keystroke data, such as keypress and release times, is identified. From this raw data, characteristics such as key hold time and key spacing are extracted to represent the unique typing patterns of individuals. Next, outlier detection is performed to identify and remove anomalous data points that may skew the analysis [3]. This step is crucial for refining the dataset and improving the accuracy of the model. Finally, classification is carried out, with the accuracy and efficiency largely depending on the specific model.

The subject of study is a mathematical model for keystroke dynamics recognition. One of the frequently employed methods in pattern recognition involves building decision rules based on prediction ellipsoids.

The purpose of the work is to increase the probability of keystroke dynamics recognition by constructing a nine-variate prediction ellipsoid for normalized data using the multivariate Box-Cox transformation (BCT).

1 PROBLEM STATEMENT

Assume we have an original data sample set consisting of nine keystroke timing features, with a multivariate distribution that is not Gaussian. In this case, there exists a bijective nine-variate normalizing transformation $\boldsymbol{\psi} = \{\psi_Y, \psi_1, \psi_2, \ldots, \psi_9\}^T$ that converts the non-Gaussian random vector $\mathbf{X} = \{X_1, X_2, \ldots, X_9\}^T$ into a Gaussian random vector $\mathbf{Z} = \{Z_1, Z_2, \ldots, Z_9\}^T$ is given by:

$$\mathbf{Z} = \boldsymbol{\psi}(\mathbf{X}). \tag{1}$$

It is required to build the prediction ellipsoid for normalized data in the form:

$$\left(\mathbf{Z} - \overline{\mathbf{Z}}\right)^T \mathbf{S}_{\mathbf{Z}}^{-1} \left(\mathbf{Z} - \overline{\mathbf{Z}}\right) = \chi^2_{m,\alpha}, \tag{2}$$

where

$$\mathbf{S}_{\mathbf{Z}} = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)\left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)^T .$$

Also, it is required to construct a prediction ellipsoid for keystroke dynamics recognition based on equation (2) and the transformations (1).

2 REVIEW OF THE LITERATURE

Mathematical modeling methods are utilized in keystroke dynamics recognition, encompassing the process of constructing and refining mathematical models to enhance accuracy and reliability. Recent advancements in this field have employed a range of techniques. Key methods include tree-based methods, such as random forests [5, 6], build hierarchical models that classify data by learning feature splits to effectively separate different classes; support vector-based approaches [7–9], establish optimal decision boundaries by maximizing the margin between classes; neural network-based models [10–12] capture complex patterns in keystroke data through multiple layers of interconnected nodes and others classifiers.

However, in developing user authentication systems, one-class classification is more typical for keystroke dynamics because it focuses on verifying whether a typing pattern belongs to a known user [13]. The primary goal is to create a model based on the unique typing patterns of an individual, and the target class, and then identify whether new typing patterns match this known profile. This is especially useful for authentication, where the system needs to continuously confirm that the current user is indeed the registered individual, rather than distinguishing between multiple users. One-class classification is connected with outlier detection [14]. The model is trained on data from a target class and does not have explicit knowledge of other classes. It identifies whether new data fits within the learned target class pattern, flagging deviations as potential outliers [15].

Popular approaches to one-class classification in keystroke dynamics recognition include one-class support vector machine [16–18], which identifies a boundary that separates target data from outliers; neural network-based classifiers such as autoencoders [19, 20], which learn to reconstruct input data and identify anomalies based on reconstruction errors, and GANs [21], which use generative adversarial networks to model target data and detect deviations. Currently, the mathematical modeling of prediction ellipsoids is widely used for pattern recognition, particularly in one-class classification systems [22, 23]. This technique employs statistical methods to define a multivariate ellipsoid that encompasses the target data points.

The construction of the ellipsoid is based on the assumption that the data follows a multivariate normal distribution. One promising solution to address this limitation is the application of normalizing transformations [24–26]. In the study [27] it was observed that the prediction ellipsoid for normalized data outperformed machine

DOI 10.15588/1607-3274-2025-1-9

learning methods such as one-class SVM, isolation forest, and autoencoder in one-class classification for face recognition tasks. We propose applying this mathematical model to keystroke dynamics recognition systems.

Normalization techniques can convert non-normal data into a distribution that more closely approximates multivariate Gaussian distribution, thereby improving the effectiveness of statistical analyses and machine learning models. While univariate normalization techniques transform each feature independently, multivariate normalization techniques consider the relationships between multiple features simultaneously. In this study, we employ the multivariate BCT to normalize keystroke dynamics data, aiming to enhance the overall recognition performance.

3 MATERIALS AND METHODS

In the context of keystroke dynamics-based recognition, the dataset used plays a pivotal role in the effectiveness and accuracy of the algorithms. A keystroke dynamics dataset typically consists of detailed records of an individual's typing behavior, capturing various temporal aspects of keystrokes such as the duration of key presses and the intervals between key presses and releases.

The CMU keystroke dynamics dataset used in this study provides a detailed record of typing from 51 subjects, each of whom typed a static password string: ".tie5Roanl". The dataset includes various keystroke timing features, recorded in seconds, such as the duration a key is held down and the time between key presses.

Data collection involved eight separate sessions per subject, with at least one day between each session. During each session, subjects typed the password 50 times, resulting in 400 repetitions per subject. This setup provides a comprehensive dataset with a total of 20,400 samples across all subjects.

The dataset is structured with a subject identifier, session number, repetition number, and a total of 31 timing features. It includes columns with specific naming conventions to represent various keystroke timing metrics. Columns labeled as H.key indicate the hold time for a particular key, capturing the duration from when the key is pressed to when it is released. DD.key1.key2 columns represent the time interval between pressing two consecutive keys, known as keydown-keydown time. Similarly, UD.key1.key2 columns denote the keyup-keydown time, which measures the interval between releasing one key and pressing the next. It is important to note that UD times can sometimes be negative, and the sum of H times and UD times corresponds to the DD time for a given digraph.

To simplify the model, it was decided to focus on 9 key properties, so the feature vector takes the form: X = { H.t, H.i, H.e, H.5, H.R, H.o, H.a, H.n, H.l }.

After extracting feature vectors, the next crucial step is outlier detection. Identifying and removing anomalies is essential because they can significantly skew the analysis and degrade the accuracy of the recognition model. By detecting outliers, the dataset is refined, ensuring that the model is trained on data that truly represents typical user behavior.

One popular method for anomaly detection is based on the squared Mahalanobis distance (SMD). A known limitation of the SMD method is its reliance on the assumption that the data follows a multivariate Gaussian distribution. To address this, it is essential to first evaluate the data's normality using statistical tests such as the Mardia test.

The Mardia test is employed to assess whether a dataset adheres to multivariate normality, which is crucial for methods assuming a multivariate Gaussian distribution. This test also applied in the study [27], evaluates two key dimensions of normality: multivariate skewness β_1 and multivariate kurtosis β_2.

Skewness measures the asymmetry of the data distribution. When skewness is scaled by $N/6$, it follows a chi-square distribution with $p(p+1)(p+2)/6$ degrees of freedom, where p is the number of variables and N is the sample size.

Kurtosis assesses the tailedness of the distribution or the extent to which the tails differ from those of a normal distribution. The Mardia test calculates kurtosis coefficients and compares them to a normal distribution with a mean of $p(p+2)$ and a variance of $8p(p+2)/N$.

If the data deviates significantly from multivariate normality, it is imperative to apply a normalizing transformation (1) on a non-Gaussian random vector $\mathbf{X} = \{X_1, X_2, \ldots, X_9\}^T$ to convert it into a Gaussian random vector $\mathbf{Z} = \{Z_1, Z_2, \ldots, Z_9\}^T$.

Normalizing transformations are essential tools in statistical analysis and machine learning, primarily because they help stabilize variance, reduce skewness, and align data more closely with a multivariate Gaussian distribution. This process is particularly critical for methods that rely on the assumption of multivariate normality, as such transformations can greatly enhance the validity and interpretability of the results. There are two main types of normalizing transformations: univariate and multivariate.

Univariate transformations include methods like the logarithm and the univariate BCT. The logarithm is commonly employed to stabilize variance, especially in data with positive skewness. On the other hand, the univariate BCT offers greater flexibility by addressing both positive and negative skewness through the selection of an optimal parameter λ. However, this method can be more complex to implement due to the need for parameter estimation, which may also be sensitive to outliers.

In contrast, multivariate transformations, such as the multivariate BCT, overcome the limitations of univariate methods by considering the interrelationships between multiple variables. While univariate transformations are more straightforward, they can fall short in situations where the interdependencies among variables are significant. Multivariate transformations provide a more

DOI 10.15588/1607-3274-2025-1-9

comprehensive approach but come with increased computational complexity.

The multivariate BCT builds on the principles of the univariate BCT approach but extends its application to multiple variables simultaneously.

$$Z_j = x(\lambda_j) = \begin{cases} \left(X_j^{\lambda_j} - 1\right) / \lambda_j, \lambda_j \neq 0; \\ \ln(X_j), \quad \lambda_j = 0. \end{cases} \tag{3}$$

This transformation maintains the correlations between variables while normalizing their distributions, making it especially effective for multivariate data. However, applying this transformation demands considerable effort due to the complexity of the required parameter estimation. A widely used approach for estimating parameters for each feature in a set involves maximizing the log-likelihood of the transformed data, as described in the study [27].

Applying the multivariate BCT can greatly improve the approximation of the data's distribution to normality. After normalization, it's crucial to run the Mardia test again to assess the transformation's success. If the test confirms that the data now follows a multivariate normal distribution, the dataset is ready for further analysis. If not, further adjustments or alternative methods may be necessary to meet the assumptions required for subsequent statistical procedures.

The next step is the construction of the prediction ellipsoid. A prediction ellipsoid is a tool in multivariate analysis used to determine whether a data point belongs to a specific target class. This method involves calculating the squared Mahalanobis distance for each point, which corresponds to the left side of the comparison. This distance is then compared against a critical value from the chi-square distribution, representing the right side of the equation:

$$\left(\mathbf{X} - \overline{\mathbf{X}}\right)^T \mathbf{S}_{\mathbf{X}}^{-1} \left(\mathbf{X} - \overline{\mathbf{X}}\right) = \chi^2_{9,\,0.005}, \tag{4}$$

where

$$\mathbf{S}_{\mathbf{X}} = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{X}_i - \overline{\mathbf{X}}\right)\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)^T .$$

The SMD follows a chi-square distribution with degrees of freedom equal to the number of features in the data, in our case is 9. This connection allows the determination of a critical value based on the desired significance level, for one-class classification tasks, a common choice is 0.005. Specifically, if a data point's SMD exceeds the critical value from the chi-square distribution, it is classified as an anomaly (an instance of another class). Conversely, if the distance is below the critical value, the data point is classified as an instance of the target class.

In cases where the data is non-normal, following the normalization process, a nine-variate prediction ellipsoid is constructed based on (4):

$$\left(\mathbf{Z} - \overline{\mathbf{Z}}\right)^T \mathbf{S}_{\mathbf{Z}}^{-1} \left(\mathbf{Z} - \overline{\mathbf{Z}}\right) = \chi^2_{9,0.005}. \tag{5}$$

The chi-square distribution quantile value is 23.59 for 9 degrees of freedom at a significance level of 0.005. If the SMD falls below the chi-square critical value, the point is considered to lie within the ellipsoid, indicating its membership in the target class.

4 EVALUATION METRICS

In one-class classification, where the goal is to distinguish between target and anomaly instances, evaluation metrics such as accuracy, specificity, precision, recall, and the F1 score play an important role in evaluating model performance [28]. These metrics are derived from classification outcomes categorized as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

In this context, true positives (TP) indicate correctly identified anomalies, while false positives (FP) represent cases mistakenly classified as anomalies or false alarms. True negatives (TN) denote correctly identified target instances, and false negatives (FN) reflect cases where actual anomalies were incorrectly classified as targets.

Accuracy measures the overall correctness of the classification, considering both target instances and anomalies: Accuracy=(TP+TN)/(TP+TN+FP+FN).

Specificity evaluates the model's capacity to correctly classify target instances. It measures the proportion of identified target instances relative to all target instances: Specificity=TN/(TN+FP).

Precision focuses on the reliability of the model when identifying anomalies, showing the proportion of true anomalies among all instances that the model classified as anomalies: Precision=TP/(TP+FP).

Recall (sensitivity) assesses the model's ability to detect all actual anomalies, it measures the proportion of true anomalies that were correctly identified out of all existing anomalies: Recall=TP/(TP+FN).

The F1 score offers a balanced evaluation by calculating the harmonic mean between precision and recall: F1 score=2*(Precision*Recall)/(Precision + Recall).

5 EXPERIMENTS

Data with the identifier s015 was randomly selected to construct the prediction ellipsoid. Meanwhile, s004 was employed in testing to verify the recognition of keystroke dynamics from a different individual.

It is essential to identify and remove outliers, which requires verifying whether the data follows a multivariate normal distribution. The Mardia test indicated that the set with the identifier s015 shows deviations from multivariate normality, as the test statistic for multivariate skewness $N\beta_1/6$ measuring 391.54, exceeds the chi-square distribution threshold of 215.53 for 165 degrees of freedom

DOI 10.15588/1607-3274-2025-1-9

at a significance level of 0.005. Similarly, the test statistic for multivariate kurtosis β_2 with a value of 113.32, surpasses the normal distribution quantile of 102.62, given a mean of 99, variance of 1.98, and 0.005 significance level. Suggesting that further normalization is needed before proceeding with analysis.

The initial step in normalization involves determining the optimal parameters. Through the application of the maximum likelihood method, the following parameter estimates were obtained: $\hat{\lambda}_1 = 0.9939$, $\hat{\lambda}_2 = 1.3605$, $\hat{\lambda}_3 = 1.2202$, $\hat{\lambda}_4 = 1.7521$, $\hat{\lambda}_5 = 2.2965$, $\hat{\lambda}_6 = 1.0447$, $\hat{\lambda}_7 = 1.6466$, $\hat{\lambda}_8 = 1.3512$, $\hat{\lambda}_9 = 2.0599$.

After applying the nine-variate BCT, the resulting set with components (3) was analyzed using the Mardia test. The test statistic for multivariate skewness $N\beta_1/6$, which is 212.07, is below the chi-square distribution's critical value of 215.53 for 165 degrees of freedom at a 0.005 significance level. However, the test statistic for multivariate kurtosis β_2, with a value of 109.01, remains above the normal distribution quantile of 102.62, given a mean of 99, variance of 1.98, and 0.005 significance level. Despite the transformation, the set still exhibits non-normality, primarily due to the influence of outliers, that distort the distribution and prevent normalization. Nevertheless, using the normalized set is preferable, as it brings the distribution closer to multivariate normality, improving the performance of the Mahalanobis distance method.

Next, the SMD is computed for each feature vector to identify potential outliers. The calculated distances are compared against the chi-square distribution critical value of 23.59 for 9 degrees of freedom at a significance level of 0.005. Any vectors exceeding this threshold are considered outliers. In this iteration, the vector with number 295 with the maximum SMD of 37.44 is removed from the set.

This iterative process continues until all outliers are removed from the set. After removing 6 outliers, the test statistic for multivariate kurtosis of the normalized set, according to the Mardia test, falls below the critical value. This highlights the significant impact of outliers on the set's distribution.

Table 1 presents the SMDs and corresponding indices for each outlier that was removed. This iterative process continued until no further significant outliers remained, resulting in a dataset that was more refined and less influenced by extreme values.

Table 1 – Removed outliers

№	SMD	Index	№	SMD	Index
1	37.44	295	6	26.963	323
2	36.962	160	7	26.868	45
3	30.742	306	8	25.776	263
4	28.833	388	9	24.515	294
5	28.662	214	10	23.972	204

As a result of the iterative outlier removal, the set with the vector of means $\overline{\mathbf{X}}$ = {0.07525; 0.07022; 0.07823; 0.063; 0.06911; 0.08829; 0.08605; 0.07505; 0.0751} was retrieved. The covariance matrix of the final set is provided in Table 2.

Table 2 – The covariance matrix of the set after removal of outliers

$0.0^{3}19$	$0.0^{4}26$	$0.0^{4}54$	$0.0^{4}13$	$0.0^{5}49$	$0.0^{4}12$	$0.0^{4}26$	$-0.0^{5}6$	$-0.0^{4}17$
$0.0^{4}26$	$0.0^{3}21$	$-0.0^{4}19$	$0.0^{5}51$	$0.0^{5}82$	$0.0^{4}14$	$0.0^{4}4$	$-0.0^{4}1$	$0.0^{5}14$
$0.0^{4}50$	$-0.0^{4}19$	$0.0^{3}25$	$0.0^{6}1$	$0.0^{5}35$	$0.0^{5}11$	$-0.0^{4}26$	$0.0^{4}27$	$0.0^{6}1$
$0.0^{4}13$	$0.0^{5}51$	$0.0^{6}1$	$0.0^{3}19$	$0.0^{4}16$	$-0.0^{5}52$	$0.0^{4}31$	$-0.0^{4}18$	$-0.0^{5}57$
$0.0^{5}49$	$0.0^{5}82$	$0.0^{5}35$	$0.0^{4}16$	$0.0^{3}13$	$0.0^{4}14$	$-0.0^{5}62$	$0.0^{5}56$	$0.0^{5}48$
$0.0^{4}12$	$0.0^{4}14$	$0.0^{5}11$	$-0.0^{5}52$	$0.0^{4}14$	$0.0^{3}12$	$-0.0^{5}36$	$0.0^{5}36$	$0.0^{5}86$
$0.0^{4}26$	$0.0^{4}4$	$-0.0^{4}26$	$0.0^{4}31$	$-0.0^{5}62$	$-0.0^{5}36$	$0.0^{3}35$	$-0.0^{4}38$	$0.0^{5}19$
$-0.0^{5}6$	$-0.0^{4}1$	$0.0^{4}27$	$-0.0^{4}18$	$0.0^{5}56$	$0.0^{5}36$	$-0.0^{4}38$	$0.0^{3}27$	$0.0^{5}7$
$-0.0^{4}17$	$0.0^{5}14$	$0.0^{6}1$	$-0.0^{5}57$	$0.0^{5}48$	$0.0^{5}86$	$0.0^{5}19$	$0.0^{5}7$	$0.0^{3}2$

The resulting sample was randomly shuffled to ensure that the data points were distributed evenly across the training and test sets, reducing any potential bias due to the order of the data. After shuffling, the dataset was split into training and test sets, each containing 50% of the data, equating to 195 vectors per set. The training set is used to construct the prediction ellipsoid, allowing the model to learn patterns and relationships from the data. The test set, on the other hand, is used to evaluate the model's performance on a distinct subset of the data that was not seen during training. This approach helps in achieving a representative evaluation of the model's performance by providing an unbiased assessment of how well the model generalizes to new, unseen data.

Table 3 displays the covariance matrix of the training set, which has a vector of means $\overline{\mathbf{X}}$ = {0.07635; 0.07052; 0.07875; 0.06254; 0.06955; 0.08806; 0.08752; 0.07447; 0.07495}.

Based on the Mardia test results, the multivariate distribution of the training sample shows deviations from multivariate normality. The test statistic for multivariate skewness $N\beta_1/6$ is 286.99, which exceeds the critical value of 215.53 from the chi-square distribution for 165 degrees of freedom at a 0.005 significance level. The test statistic for multivariate kurtosis β_2 is 105.43, which exceeds the critical value of 104.19 for a normal distribution with a mean of 99, a variance of 4.062, and a significance level of 0.005.

DOI 10.15588/1607-3274-2025-1-9

Table 3 – The covariance matrix of the training set

$0.0^{3}2$	$0.0^{4}16$	$0.0^{4}62$	$0.0^{4}42$	$0.0^{5}53$	$0.0^{5}71$	$0.0^{4}38$	$0.0^{5}14$	$-0.0^{5}75$
$0.0^{4}16$	$0.0^{3}21$	$-0.0^{4}19$	$-0.0^{5}26$	$-0.0^{6}35$	$0.0^{4}14$	$0.0^{4}62$	$-0.0^{4}1$	$0.0^{5}15$
$0.0^{4}62$	$-0.0^{4}19$	$0.0^{3}26$	$-0.0^{6}84$	$0.0^{4}11$	$0.0^{5}71$	$-0.0^{4}51$	$0.0^{5}99$	$0.0^{4}17$
$0.0^{4}42$	$-0.0^{5}26$	$-0.0^{6}84$	$0.0^{3}21$	$0.0^{4}24$	$-0.0^{6}75$	$0.0^{4}37$	$-0.0^{4}28$	$-0.0^{5}87$
$0.0^{5}53$	$-0.0^{6}35$	$0.0^{4}11$	$0.0^{4}24$	$0.0^{3}13$	$0.0^{4}12$	$-0.0^{5}36$	$-0.0^{5}53$	$0.0^{5}72$
$0.0^{5}71$	$0.0^{4}14$	$0.0^{5}71$	$-0.0^{6}75$	$0.0^{4}12$	$0.0^{3}12$	$0.0^{4}1$	$0.0^{5}16$	$0.0^{4}18$
$0.0^{4}38$	$0.0^{4}62$	$-0.0^{4}51$	$0.0^{4}37$	$-0.0^{5}36$	$0.0^{4}1$	$0.0^{3}36$	$-0.0^{4}61$	$0.0^{6}69$
$0.0^{5}14$	$-0.0^{4}1$	$0.0^{5}99$	$-0.0^{4}28$	$-0.0^{5}53$	$0.0^{5}16$	$-0.0^{4}61$	$0.0^{3}27$	$0.0^{4}23$
$-0.0^{5}74$	$0.0^{5}15$	$0.0^{4}17$	$-0.0^{5}87$	$0.0^{5}72$	$0.0^{4}18$	$0.0^{6}69$	$0.0^{4}23$	$0.0^{3}22$

The training set is normalized using a nine-variate BCT. Optimal parameters for this transformation are determined through the maximum likelihood method: $\hat{\lambda}_1 = 1.3676$, $\hat{\lambda}_2 = 1.4807$, $\hat{\lambda}_3 = 1.078$, $\hat{\lambda}_4 = 1.7393$, $\hat{\lambda}_5 = 2.1004$, $\hat{\lambda}_6 = 1.1498$, $\hat{\lambda}_7 = 1.566$, $\hat{\lambda}_8 = 1.1685$, $\hat{\lambda}_9 = 2.1146$.

As a result of applying the BCT to the training set with components (3), the resulting sample has a vector of means: $\overline{\mathbf{Z}}$ = {0.70932; 0.66184; 0.86764; –0.57016; –0.47427; 0.81642; 0.62417; –0.81443; –0.47084}. The covariance matrix of the normalized sample is presented in Table 4.

Table 4 – The covariance matrix of the normalized set

$0.0^{4}29$	$0.0^{5}14$	$0.0^{4}2$	$0.0^{5}2$	$0.0^{6}13$	$0.0^{5}21$	$0.0^{5}39$	$0.0^{7}73$	$-0.0^{7}68$
$0.0^{5}14$	$0.0^{4}16$	$-0.0^{5}48$	$-0.0^{7}21$	$0.0^{7}71$	$0.0^{5}27$	$0.0^{5}43$	$-0.0^{5}16$	$0.0^{7}31$
$0.0^{4}2$	$-0.0^{5}48$	$0.0^{3}18$	$-0.0^{6}42$	$0.0^{6}43$	$0.0^{5}42$	$-0.0^{4}11$	$0.0^{5}51$	$0.0^{6}65$
$0.0^{5}2$	$-0.0^{7}21$	$-0.0^{6}42$	$0.0^{5}31$	$0.0^{6}15$	$-0.0^{6}15$	$0.0^{5}14$	$-0.0^{5}2$	$-0.0^{7}32$
$0.0^{6}13$	$0.0^{7}71$	$0.0^{6}43$	$0.0^{6}15$	$0.0^{6}33$	$0.0^{6}58$	$0.0^{8}8$	$-0.0^{6}17$	$0.0^{6}13$
$0.0^{5}21$	$0.0^{5}27$	$0.0^{5}42$	$-0.0^{6}15$	$0.0^{6}58$	$0.0^{4}55$	$0.0^{5}15$	$0.0^{6}92$	$0.0^{6}72$
$0.0^{5}39$	$0.0^{5}43$	$-0.0^{4}11$	$0.0^{5}14$	$0.0^{8}8$	$0.0^{5}15$	$0.0^{4}22$	$-0.0^{4}1$	$0.0^{7}12$
$0.0^{7}73$	$-0.0^{5}16$	$0.0^{5}51$	$-0.0^{5}2$	$-0.0^{6}17$	$0.0^{6}92$	$-0.0^{4}1$	$0.0^{3}11$	$0.0^{6}89$
$-0.0^{7}68$	$0.0^{7}31$	$0.0^{6}66$	$-0.0^{7}32$	$0.0^{7}13$	$0.0^{6}72$	$0.0^{7}12$	$0.0^{6}89$	$0.0^{6}57$

According to the Mardia test, the normalized training set conforms to the multivariate normal distribution. The test statistic for multivariate skewness $N\beta1/6$ is 175.47, which is below the critical value of 215.53 from the chi-square distribution with 165 degrees of freedom at a 0.005 significance level. Additionally, the test statistic for multivariate kurtosis $\beta2$ is 99.76, which does not exceed the critical value of 104.19 for a normal distribution with a mean of 99, a variance of 4.062, and a significance level of 0.005.

6 RESULTS

After applying the normalization, nine-variate prediction ellipsoids were constructed for both the non-Gaussian data (NGD) (4) and the normalized data (ND) (5). The computer program implementing the constructed models was developed to conduct experiments. The program was written in the Python language.

Table 5 presents a comparative analysis of the performance metrics for prediction ellipsoid models in one-class classification.

Table 5 – The covariance matrix of the training set

Model	Accuracy	Specificity	Precision	Recall	F1 score
NGD	0.9412	0.9795	0.9893	0.9225	0.9547
ND	0.9782	0.9949	0.9974	0.9700	0.9835

The metrics for the nine-variate prediction ellipsoids for non-Gaussian data reflect good performance, which is likely attributed to the effective keystroke dynamics recognition. However, the nine-variate prediction ellipsoid for normalized data significantly outperforms it. Normalization has led to better detection of anomalies, reduced false positives, and improved overall classification performance. The results highlight the importance of applying normalization techniques to achieve a more reliable and accurate model for keystroke dynamics recognition.

7 DISCUSSION

The results demonstrate that applying the nine-variate BCT significantly improved the model's performance, underscoring the importance of normalization in addressing non-Gaussian data distributions. Choosing the appropriate normalization technique is vital, as it significantly affects the model's effectiveness and reliability. Multivariate normalization methods capture complex variable relationships, enhancing prediction ellipsoid accuracy and providing a more precise representation of data patterns.

The choice of significance level is an important factor in constructing the prediction ellipsoid. In this study, a

DOI 10.15588/1607-3274-2025-1-9

significance level of 0.005 was selected, aligning with common practices in one-class classification and outlier detection tasks [29].

Despite these advantages, using the prediction ellipsoid for normalized data comes with some disadvantages. It is generally considered essential to have at least 100 instances for building a high-quality model. Additionally, selecting the appropriate normalization transformation can be challenging, particularly for sets that have complex distributions or many outliers. The last is the need to choose a significance level that affects the efficiency and reliability of the prediction ellipsoid.

Since 10 data points were removed as outliers, the model might miss some underlying patterns. This limitation might be addressed with a more complex normalization, such as the Johnson transformation, which could better handle the nuances of the data distribution and improve the model's ability to represent all relevant data points.

CONCLUSIONS

The study examined the influence of data distribution on keystroke dynamics recognition, with a particular focus on the application of a nine-variate prediction ellipsoid for normalized data using the multivariate BCT.

The research evaluated the transformation's impact on performance by comparing it with a prediction ellipsoid model developed for non-Gaussian data. The findings demonstrated that applying the BCT significantly improved model performance. Normalizing the data led to a more accurate and robust prediction ellipsoid, outperforming the non-normalized model across various evaluation metrics. The nine-variate BCT not only enhanced overall model performance but also offered deeper insights into feature relationships by accounting for correlations. This underscores the critical importance of selecting an appropriate normalization technique when dealing with non-Gaussian distributions.

Despite the benefits of normalization, the study also identified disadvantages, particularly in determining the optimal normalization transformation. Additionally, selecting an appropriate significance level is important, as it directly impacts the reliability and effectiveness of the prediction ellipsoid.

The scientific novelty of the obtained results is that the nine-variate prediction ellipsoid for normalized data for solving keystroke dynamics recognition tasks is first built using the nine-variate BCT. The application of a constructed prediction ellipsoid for normalized data allowed to increase accuracy.

The practical significance of the obtained results is that the software implementing the constructed ellipsoid is developed in the Python language. The experimental results allow us to recommend the constructed model for use in practice.

Prospects for further research could explore alternative normalization techniques, such as the Johnson transformation, to further refine model performance. Additionally, investigating the effects of model complexity and feature selection could provide valuable insights into improving methods for keystroke dynamics recognition.

ACKNOWLEDGEMENTS

The work is of an initiative nature and was carried out within the scope of the authors' scientific activities during their regular working hours in their primary positions.

REFERENCES

1. Dias T., Vitorino J., Maia E., Sousa O., Praça I. KeyRecs: A keystroke dynamics and typing pattern recognition dataset, *Data in Brief,* 2023, Vol. 50, pp. 1–8. DOI: 10.1016/j.dib.2023.109509
2. Alshehri A., Coenen F., Bollegala D. Accurate continuous and non-intrusive user authentication with multivariate keystroke streaming, *9th International Conference on Knowledge Discovery and Information Retrieval, Funchal, 1–3 November, 2017 : proceeding*s. Funchal, SciTePress, 2017, pp. 61–70. DOI: 10.5220/0006497200610070
3. Ismail M., Salem M., Abd El Ghany M.et al. Outlier detection for keystroke biometric user authentication, *PeerJ Computer Science,* 2024, Vol. 50. DOI: 10.7717/peerj-cs.2086
4. Oyebola O. Examining the distribution of keystroke dynamics features on computer, tablet and mobile phone platforms, *Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2023, Nepal, 11–12 January, 2023 : proceedings*. Singapore, Springer Nature Singapore 2023, pp. 613–620 DOI: 10.1007/978-981-99-0835-6_43
5. Saini B., Singh P., Nayyar A. , et al. A three-step authentication model for mobile phone user using keystroke dynamics, *IEEE Access*, 2020, Vol. 10, pp. 125909–125922. DOI: 10.1109/ACCESS.2020.3008019
6. Anusas-Amornkul N. Strengthening password authentication using keystroke dynamics and smartphone sensors, *ICICM 2019: Proceedings of the 9th International Conference on Information Communication and Management, Prague, 23–26 August, 2019 : proceedings*. New York, Association for Computing Machinery, 2019, pp. 70–74. DOI: 10.1145/3357419.3357425
7. Fierrez J., Pozo A., Martinez-Diaz M. et al. Benchmarking touchscreen biometrics for mobile authentication, *IEEE Transactions on Information Forensics and Security,* 2018. Vol. 13, № 11, pp. 2720–2733. DOI: 10.1109/TIFS.2018.2833042
8. Li Q. Chen H. CDAS: A continuous dynamic authentication system, *Proceedings of the 2019 8th International Conference on Software and Computer Applications, Penang, 19–21 February, 2019, proceedings.* New York, Association for Computing Machinery, 2019, pp. 447–452. DOI: 10.1145/3316615.3316691
9. Chen L., Zhong Y., Ai W. et al. Continuous authentication based on user interaction behavior / [L. Chen,] // *2019 7th International Symposium on Digital Forensics and Security (ISDFS), Barcelos, 10–12 June, 2019 : proceedings*. Barcelos, IEEE, 2019, pp. 1–6. DOI: 10.1109/ISDFS.2019.8757539
10. Baynath P., Soyjaudah K., Khan M. Keystroke recognition using neural network, *2017 5th International Sympo-*

DOI 10.15588/1607-3274-2025-1-9

sium on Computational and Business Intelligence (ISCBI), Dubai, 11–14 August, 2017 : proceedings. Dubai, IEEE, 2017, pp. 86–90. DOI: 10.1109/ISCBI.2017.8053550

11. Andrean A., Jayabalan M., Thiruchelvam V. Keystroke dynamics based user authentication using deep multilayer perceptron, *International Journal of Machine Learning and Computing,* 2020, Vol. 10, № 1, pp. 134–139. DOI: 10.18178/ijmlc.2020.10.1.910
12. Sharma A., Jureček M., Stamp M. Keystroke dynamics for user identification, *arXiv preprint arXiv:2307.05529,* 2023, pp. 1–22. DOI: 10.48550/arXiv.2307.05529
13. Raul N., Shankarmani R., Joshi P. A comprehensive review of keystroke dynamics-based authentication mechanism, *International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, Delhi, 20–21 February, 2021 : proceedings*. Singapore, Springer, 2021, Vol. 1059, pp.149–162 DOI: 10.1007/978-981-15-03245_13
14. Antal M., Szabó L. An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices, *2015 20th International Conference on Control Systems and Computer Science, Bucharest, 27–29 May, 2015 : proceedings*. Bucharest, IEEE, 2015, pp. 343–350. DOI: 10.1109/CSCS.2015.16
15. Marques H., Swersky L., Sander J. et al. On the evaluation of outlier detection and one-class classification: a comparative study of algorithms, model selection, and ensembles, *Data Mining and Knowledge Discovery,* 2023, Vol. 37, pp. 1–45. DOI: 10.1007/s10618-023-00931-x
16. Toosi R., Akhaee M. Time-frequency analysis of keystroke dynamics for user authentication, *Future Generation Computer Systems,* 2021, Vol. 115, pp. 438–447. DOI: 10.1016/j.future.2020.09.027
17. Jawed H., Ziad Z., Khan M. et al. Anomaly detection through keystroke and tap dynamics implemented via machine learning algorithms, *Turkish Journal of Electrical Engineering and Computer Sciences,* 2018, Vol. 26, № 4, pp. 1698–1709. DOI: 10.3906/elk-1711-410
18. Ceker H., Upadhyaya S. User authentication with keystroke dynamics in long-text data, *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, 6–9 September, 2016*. Niagara Falls, IEEE, 2016, pp. 1–6. DOI: 10.1109/BTAS.2016.7791182
19. Patel Y., Ouazzane K., Vassilev V. et al. Keystroke dynamics using auto encoders, *2019 International Conference on Cyber Security and Protection of Digital Services, Oxford, 3–4 June, 2019 : proceedings*. Oxford, IEEE, 2019, pp. 1–8. DOI: 10.1109/CyberSecPODS.2019.8885203
20. Centeno M., Moorsel A., Castruccio S. Smartphone continuous authentication using deep learning autoencoders, *15th Annual Conference on Privacy, Security and Trust (PST), Calgary, 28–30 August, 2017 : proceedings.* Calgary, IEEE, 2017, pp. 1470–1478. DOI: 10.1109/PST.2017.00026
21. Eizaguirre-Peral I., Segurola-Gil L., Zola F. Conditional generative adversarial network for keystroke presentation attack, *arXiv preprint arXiv:2212.08445,* 2022, pp. 1–4. DOI: 10.48550/arXiv.2212.08445
22. Kim S., Park D., Jung J. Evaluation of one-class classifiers for fault detection: Mahalanobis classifiers and the Mahalanobis-Taguchi system, *Processes,* 2021, Vol. 9, № 8, P. 1450. DOI: 10.3390/pr9081450
23. Bezerra V., Da Costa V., Barbon Junior S.et al. IoTDS: A one-class classification approach to detect botnets in Internet of Things devices, *Sensors,* 2019, Vol. 19, № 14, P. 3188. DOI: 10.3390/s19143188
24. Prykhodko S. B., Shutko I. S., Prykhodko A. S. A nonlinear regression model to estimate the size of web apps created using the CakePHP framework, *Radio Electronics, Computer Science, Control*, 2021, Vol. 59, № 4, pp. 129–139. DOI: 10.15588/1607-3274-2021-4-12
25. Lam K., Meijer K., Loonstra F. et al. Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis, *Multiple Sclerosis Journal,* 2021, Vol. 27, № 4, pp. 1421–1431. DOI: 10.1177/1352458520968797
26. Prykhodko S., Trukhov A. Face recognition using the ten-variate prediction ellipsoid for normalized data based on the Box-Cox transformation, *Radio Electronics, Computer Science, Control,* 2024, Vol. 2, pp. 82–89. DOI: 10.15588/1607-3274-2024-2-9
27. Prykhodko S., Trukhov A. Application of a ten-variate prediction ellipsoid for normalized data and machine learning algorithms for face recognition, *CMIS-2024: Seventh International Workshop on Computer Modeling and Intelligent Systems, Zaporizhzhia, 3 May, 2024 : proceedings*. Aachen, CEUR, 2024, pp. 362–375.
28. Nurmaini S., Darmawahyuni A., Sakti Mukti A.et al. Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification, *Electronics,* 2020, Vol. 9, № 1, P. 135. DOI: 10.3390/electronics9010135
29. Prykhodko S., Prykhodko N., Makarova L. et al. Detecting outliers in multivariate non-Gaussian data on the basis of normalizing transformations, *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering, Kyiv, 29 May – 2 June, 2017 : proceedings*. New York, IEEE, 2017, pp. 846–849. DOI: 10.1109/UKRCON.2017.8100366

Received 11.11.2024.
Accepted 10.01.2025.

DOI 10.15588/1607-3274-2025-1-9

УДК 004.93

РОЗПІЗНАВАННЯ КЛАВІАТУРНОГО ПОЧЕРКУ ЗА ДОПОМОГОЮ ДЕВ'ЯТИВИМІРНОГО ЕЛІПСОЇДА ПРОГНОЗУВАННЯ ДЛЯ НОРМАЛІЗОВАНИХ ДАНИХ

Приходько С. Б. – д-р техн. наук, професор, завідувач кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування ім. адмірала Макарова, Миколаїв, Україна.

Трухов А. С. – аспірант кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування ім. адмірала Макарова, Миколаїв, Україна.

АНОТАЦІЯ

Актуальність. Розпізнавання клавіатурного почерку є важливим елементом у підвищенні безпеки, що дозволяє реалізувати персоналізовану автентифікацію користувачів та підтримує різні системи перевірки особистості. Це дослідження вивчає вплив розподілу даних на ефективність моделей однокласової класифікації в задачах розпізнавання клавіатурного почерку, зосереджуючи увагу на застосуванні дев'ятивимірного еліпсоїда прогнозування. Об'єктом дослідження є процес розпізнавання клавіатурного почерку. Предметом дослідження є математичні моделі для розпізнавання клавіатурного почерку. На відміну від типових підходів, що передбачають багатовимірний нормальний розподіл даних, реальні набори даних часто відхиляється від нього, що ускладнює побудову точних і надійних моделей. Для вирішення цієї проблеми дані були нормалізовані за допомогою багатовимірного перетворення Бокса-Кокса, що дозволило покращити вірогідність розпізнавання клавіатурного почерку за допомогою застосування еліпсоїда прогнозування для нормалізованих даних.

Метою роботи є підвищення ймовірності розпізнавання клавіатурного почерку шляхом побудови дев'ятивимірного еліпсоїда прогнозування для нормалізованих даних із використанням багатовимірного перетворення Бокса-Кокса.

Метод. Дослідження включає побудову дев'ятивимірного еліпсоїда прогнозування для даних, нормалізованих за допомогою перетворення Бокса-Кокса. Квадрат відстані Махаланобіса застосовується для виявлення та видалення викидів, а тест Мардіа оцінює відхилення багатовимірного розподілу від нормального. Оцінки параметрів багатовимірного перетворення Бокса-Кокса отримані методом максимальної правдоподібності.

Результати. Результати показують значне підвищення вірогідності розпізнавання після нормалізації, що полягає у збільшені точності та надійності порівняно з моделями, побудованими для ненормалізованих даних. Застосування дев'ятивимірного перетворення Бокса-Кокса дозволило краще врахувати кореляції між ознаками, що дозволило еліпсоїду прогнозування краще захоплювати складні закономірності даних.

Висновки. Для розпізнавання клавіатурного почерку була розроблена математична модель у формі дев'ятивимірного еліпсоїда прогнозування для даних, нормалізованих із використанням багатовимірного перетворення Бокса-Кокса, що підвищує ймовірність розпізнавання в порівнянні з моделями, побудованими для ненормалізованих даних. Однак залишаються труднощі у визначенні оптимального методу нормалізації та виборі рівня значущості для побудови еліпсоїда прогнозування. Ці висновки підкреслюють важливість ретельного вибору ознак та застосування вдосконалених методів нормалізації даних для подальших досліджень у сфері розпізнавання клавіатурного почерку.

КЛЮЧОВІ СЛОВА: розпізнавання клавіатурного почерку, багатовимірне перетворення Бокса–Кокса, еліпсоїд прогнозування, нормалізуюче перетворення.

ЛІТЕРАТУРА

1. KeyRecs: A keystroke dynamics and typing pattern recognition dataset / [T. Dias, J. Vitorino, E. Maia et al.] // Data in Brief. – 2023. – Vol. 50. – P. 1–8. DOI: 10.1016/j.dib.2023.109509
2. Alshehri A. Accurate continuous and non-intrusive user authentication with multivariate keystroke streaming / A. Alshehri, F. Coenen, D. Bollegala // 9th International Conference on Knowledge Discovery and Information Retrieval, Funchal, 1–3 November, 2017 : proceedings. – Funchal : SciTePress, 2017. – P. 61–70. DOI: 10.5220/0006497200610070
3. Ismail M. Outlier detection for keystroke biometric user authentication / [M. Ismail, M. Salem, M. Abd El Ghany et al.] // PeerJ Computer Science. – 2024. – Vol. 50. DOI: 10.7717/peerj-cs.2086
4. Oyebola O. Examining the distribution of keystroke dynamics features on computer, tablet and mobile phone platforms / O. Oyebola // Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2023, Nepal, 11–12 January, 2023 : proceedings. – Singapore: Springer Nature Singapore 2023. – P. 613–620 DOI: 10.1007/978-981-99-0835-6_43
5. A three-step authentication model for mobile phone user using keystroke dynamics / [B. Saini, P. Singh, A. Nayyar, et al.] // IEEE Access. – 2020. – Vol. 10. – P. 125909–125922. DOI: 10.1109/ACCESS.2020.3008019
6. Anusas-Amornkul N. Strengthening password authentication using keystroke dynamics and smartphone sensors / N. Anusas-Amornkul // ICICM 2019: Proceedings of the 9th International Conference on Information Communication and Management, Prague, 23 – 26 August, 2019 : proceedings. – New York : Association for Computing Machinery, 2019. – P. 70–74. DOI: 10.1145/3357419.3357425
7. Benchmarking touchscreen biometrics for mobile authentication / [J. Fierrez, A. Pozo, M. Martinez-Diaz et al.] // IEEE Transactions on Information Forensics and Security. – 2018. – Vol. 13, № 11. – P. 2720–2733. DOI: 10.1109/TIFS.2018.2833042
8. Li Q. CDAS: A continuous dynamic authentication system / Q. Li, H. Chen // Proceedings of the 2019 8th International Conference on Software and Computer Applications, Penang, 19–21 February, 2019 : proceedings. – New York : Association for Computing Machinery, 2019. – P. 447–452. DOI: 10.1145/3316615.3316691
9. Continuous authentication based on user interaction behavior / [L. Chen, Y. Zhong, W. Ai et al.] // 2019 7th International Symposium on Digital Forensics and Security (ISDFS), Barcelos, 10–12 June, 2019 : proceedings. – Barcelos : IEEE, 2019. – P. 1–6. DOI: 10.1109/ISDFS.2019.8757539

DOI 10.15588/1607-3274-2025-1-9

10. Baynath P. Keystroke recognition using neural network / P. Baynath, K. Soyjaudah, M. Khan // 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), Dubai, 11–14 August, 2017 : proceedings. – Dubai : IEEE, 2017. – P. 86–90. DOI: 10.1109/ISCBI.2017.8053550
11. Andrean A. Keystroke dynamics based user authentication using deep multilayer perceptron / A. Andrean, M. Jayabalan, V. Thiruchelvam // International Journal of Machine Learning and Computing. – 2020. – Vol. 10, № 1. – P. 134–139. DOI: 10.18178/ijmlc.2020.10.1.910
12. Sharma A. Keystroke dynamics for user identification / A. Sharma, M. Jureček, M. Stamp // arXiv preprint arXiv:2307.05529, 2023. – P. 1–22. DOI: 10.48550/arXiv.2307.05529
13. Raul N. A comprehensive review of keystroke dynamics-based authentication mechanism / N. Raul, R. Shankarmani, P. Joshi // International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, Delhi, 20–21 February, 2021 : proceedings. – Singapore : Springer, 2021. – Vol. 1059. – P.149–162 DOI: 10.1007/978-981-15-03245_13
14. Antal M. An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices / M. Antal, L. Szabó // 2015 20th International Conference on Control Systems and Computer Science, Bucharest, 27–29 May, 2015 : proceedings. Bucharest : IEEE, 2015. – P. 343–350. DOI: 10.1109/CSCS.2015.16
15. On the evaluation of outlier detection and one-class classification: a comparative study of algorithms, model selection, and ensembles / [H. Marques, L. Swersky, J. Sander et al.] // Data Mining and Knowledge Discovery. – 2023. – Vol. 37. – P. 1–45. DOI: 10.1007/s10618-023-00931-x
16. Toosi R. Time-frequency analysis of keystroke dynamics for user authentication / R. Toosi, M. Akhaee // Future Generation Computer Systems. – 2021. – Vol. 115 – P. 438–447. DOI: 10.1016/j.future.2020.09.027
17. Jawed H. Anomaly detection through keystroke and tap dynamics implemented via machine learning algorithms / [H. Jawed, Z. Ziad, M. Khan et al.] // Turkish Journal of Electrical Engineering and Computer Sciences. – 2018. – Vol. 26, № 4. – P. 1698–1709. DOI: 10.3906/elk-1711-410
18. Ceker H. User authentication with keystroke dynamics in long-text data / H. Ceker, S. Upadhyaya // 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, 6–9 September, 2016. – Niagara Falls : IEEE, 2016. – P. 1–6. DOI: 10.1109/BTAS.2016.7791182
19. Keystroke dynamics using auto encoders / [Y. Patel, K. Ouazzane, V. Vassilev et al.] // 2019 International Conference on Cyber Security and Protection of Digital Services, Oxford, 3–4 June, 2019 : proceedings. – Oxford : IEEE, 2019. – P. 1–8. DOI: 10.1109/CyberSecPODS.2019.8885203
20. Centeno M. Smartphone continuous authentication using deep learning autoencoders / M. Centeno, A. Moorsel, S. Castruccio // 15th Annual Conference on Privacy, Security and Trust (PST), Calgary, 28–30 August, 2017 : proceedings. – Calgary : IEEE, 2017. – P. 1470–1478. DOI: 10.1109/PST.2017.00026
21. Eizaguirre-Peral I. Conditional generative adversarial network for keystroke presentation attack / I. Eizaguirre-Peral, L. Segurola-Gil, F. Zola // arXiv preprint arXiv:2212.08445 2022. – P. 1–4. DOI: 10.48550/arXiv.2212.08445
22. Kim S. Evaluation of one-class classifiers for fault detection: Mahalanobis classifiers and the Mahalanobis–Taguchi system / S. Kim, D. Park, J. Jung // Processes. – 2021. – Vol. 9, № 8. – P. 1450. DOI: 10.3390/pr9081450
23. IoTDS: A one-class classification approach to detect botnets in Internet of Things devices / [V. Bezerra, V. da Costa, S. Barbon Junior et al.] // Sensors. – 2019. – Vol. 19, № 14. – P. 3188. DOI: 10.3390/s19143188
24. Prykhodko S. B. A nonlinear regression model to estimate the size of web apps created using the CakePHP framework / S. B. Prykhodko, I. S. Shutko, A. S. Prykhodko // Radio Electronics, Computer Science, Control. – 2021. – Vol. 59, № 4. – P. 129–139. DOI: 10.15588/1607-3274-2021-4-12
25. Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis / [K. Lam, K. Meijer, F. Loonstra et al.] // Multiple Sclerosis Journal. – 2021. – Vol. 27, № 4. – P. 1421–1431. DOI: 10.1177/1352458520968797
26. Prykhodko S. Face recognition using the ten-variate prediction ellipsoid for normalized data based on the Box-Cox transformation / S. Prykhodko, A. Trukhov // Radio Electronics, Computer Science, Control. – 2024. – Vol. 2. – P. 82–89. DOI: 10.15588/1607-3274-2024-2-9
27. Prykhodko S. Application of a ten-variate prediction ellipsoid for normalized data and machine learning algorithms for face recognition / S. Prykhodko, A. Trukhov // CMIS-2024: Seventh International Workshop on Computer Modeling and Intelligent Systems, Zaporizhzhia, 3 May, 2024 : proceedings. – Aachen : CEUR, 2024. – P. 362–375.
28. Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification / [S. Nurmaini, A. Darmawahyuni, A. Sakti Mukti et al.] // Electronics. – 2020. – Vol. 9, № 1. – P. 135. DOI: 10.3390/electronics9010135
29. Detecting outliers in multivariate non-Gaussian data on the basis of normalizing transformations / [S. Prykhodko, N. Prykhodko, L. Makarova et al.] // 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering, Kyiv, 29 May – 2 June, 2017 : proceedings. – New York : IEEE, 2017. – P. 846–849. DOI: 10.1109/UKRCON.2017.8100366

DOI 10.15588/1607-3274-2025-1-9