Keystroke Dynamics Based Authentication System

Nurşah Çevik
Research and Development Department
IBSS Consulting
İstanbul, Turkey
nursah.cevik@ibss.com.tr

Sedat Akleylek
Department of Computer Engineering
Ondokuz Mayıs University
Samsun, Turkey
sedat.akleylek@bil.omu.edu.tr

Kadir Yunus Koç
Research and Development Department
IBSS Consulting
İstanbul, Turkey
kadiryunus.koc@ibss.com.tr

2021 6th International Conference on Computer Science and Engineering (UBMK) | 978-1-6654-2908-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/UBMK52708.2021.9559008

***Abstract*—Nowadays, many companies use biometric technologies for the security of critical systems as well as username-password methods. In the literature, biometric systems are the most commonly used systems among the two-factor authentication systems. There are two different approaches to biometric systems: physical and behavioral biometric systems. In the last decade, the accuracy of behavioral biometric systems has significantly increased with the use of machine learning methods in these systems. For this reason, the usage areas of the studies in this field have expanded. In this study, we focus on keystroke dynamics based on behavioral methods. Firstly, we make a web application to collect keystroke data from 54 employees in a company. Then, we use the benchmark database and our database to train-test machine learning algorithms, which have the highest accuracy in this field in the literature. Among them, tree-based algorithms have the highest accuracy score, with an average of 0.94.**

***Index Terms*—keystroke dynamics, behavioral biometrics, two-factor authentication, machine learning**

I. Introduction

The authentication systems provide that communicating parties can authenticate each other's identity. In this process, something you are, something you have, or something you know is used to confirm your identity. Today, in most systems, we use only something you know, such as username and password, in the authentication process. The security of these systems relies on one-factor authentication [1].

Today, in most systems, only the username and password are used in the authentication process, and this method is called one-factor authentication. However, this method does not provide the desired security level due to user errors such as easy password selection, the capture of the password by a malicious person, the sharing of the password with unauthorized persons, and the use of similar passwords in different systems. For this reason, it is recommended to use the multi-factor authentication mechanism to ensure the desired level of security and reduce the possibility of unauthorized access [2]. In the multi-factor authentication mechanism, users must have two different proofs to verify their identity. For example, they can authenticate using a fingerprint or/and a smart card besides a username and password.

Nowadays, in addition to the fact that multi-factor authentication is mandatory in critical systems, it is recommended to use it in many non-critical areas for security. Authentication with Short Message Services (SMS) is one of the most widely used techniques among multi-factor authentication mechanisms in the literature [3]. According to the latest legislation published by the Banking Regulation and Supervision Agency (Bankacılık Düzenleme ve Denetleme Kurumu-BDDK) in Turkey [4], the SMS verification is not suitable for the second factor in the multi-factor authentication process. Furthermore, it requires an extra operation for the user and includes a cost per message sent. Consequently, it is seen that SMS authentication will not be preferred in the future, and so different approaches will be needed instead of this technique [5]. In the literature, using biometrics is the most common technique to provide multi-factor authentication. This method is divided into two different parts: physical biometric methods like fingerprints, retina, and behavioral biometric methods like the pattern of gait, keystroke dynamics, mouse dynamics. Within the scope of this study, studies on behavioral biometric methods are detailed examined. It has been observed that these techniques do not require additional hardware and do not affect the user experience while controlling user identity. Therefore, they can easily integrate into the system. For this reason, behavioral biometrics-based authentication systems are currently used in the security of many critical systems in the public and private sectors.

In the last decade, among behavioral biometric methods, the accuracy of techniques used in keystroke dynamics has increased with advances in artificial intelligence and machine learning [6]. Therefore, it has become a viable alternative among other biometric features. Pin Shen Teh et al. in 2013 [7], examined the studies between 1980-2012 and explained in detail the transition from classical statistical algorithms to machine learning algorithms in this field. In the study, the systems are compared according to their data collection methods, classification algorithms, feature extraction methods, and protocols. Later, they stated that besides the accuracy, features such as efficiency, adaptation to change, and ease of use should be evaluated as success criteria for these systems. As a result, they stated that keystroke dynamics are suitable for two-factor authentication systems. Ali et al. compared the approaches used in studies presented between 1980 and 2014 from different aspects and showed the effect of data collection devices' algorithms on the system performance [8], [9]. Also, they stated that the running time and efficiency of the systems are significant constraints for systems since the authentication process is a real-time process [8]. Pahuja and Nagabhushan

978-1-6654-2908-5/21/$31.00 ©2021 IEEE

addressed the two main problems of authentication systems based on keystroke dynamics: first, human behavior changes over time; second, there is no standardized protocol for evaluating systems [10]. This study focuses on authentication systems based on keystroke dynamics.

A. Motivation and Our Contribution

The literature review shows that two-factor authentication systems will become obligatory for many systems in the future. Authentication systems that use keystroke dynamics were proved to be a powerful alternative between two-factor authentication systems, thanks to their high accuracy and user-friendly nature. For this reason, we design an authentication system based on the keystroke dynamics, taking into account the success criteria and algorithms used in previous studies. In this system, models are established with different classification algorithms using keyboard data obtained from 54 people. Also, we classify the benchmark dataset used in the study presented in [11] with establish models and share the classification results in detail.

B. Organization

The organization of this paper is as follows: In Section II, we describe material and methods for keystroke dynamics based authentication systems in the literature. In Section III, we define our data collection application and processes. After that, we explain the classification algorithms we use in our system in detail and visualize the results of the algorithms. Finally, we compare the results with the previous studies in the literature and present the limits of these systems. In Section IV, the conclusion and future studies are given.

II. MATERIALS AND METHODS

In this section, we define the metric types of keystroke dynamics in the literature. Then, we give general information about the data collection step of these studies and explain the most common methods used to authenticate a person's identity using keystroke dynamics.

A. Keystroke Dynamics

In the literature, it was observed that four different time metrics were taken into account while collecting the keystroke dynamics. These time metrics are shown in Figure 1.

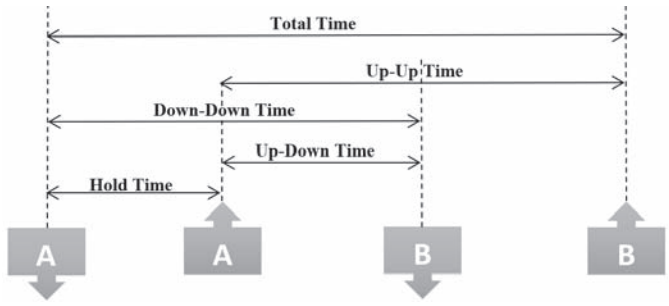

Fig. 1. Time Metrics.

1) **Hold Time:** The time between pressing a key and releasing that key.
2) **Up-Down Time (UD Time):** The time between releasing a key and pressing the next key.
3) **Down-Down Time (DD Time):** The time between pressing a key and pressing the next key.
4) **Up-Up Time (UU Time):** The time between releasing a key and releasing the next key.

In this study, the user's time to click on keyboard keys and the time to transition between keys are calculated, and behavior patterns are obtained for 54 users. Since the first three metrics are the most commonly used to identify users, these metrics were used when calculating users' keyboard patterns. Other contextual data, such as age, gender, and hand, will be used as a feature in future studies.

B. Data Collection

In the literature, the data were generally collected from university students. Therefore, the patterns in these databases represent the typing characters of people in a similar demographic group [11]. That poses a problem in the train and test phases of algorithms. For this reason, we collect data from a company employees with different demographic characteristics from the university students. Therefore, the results of this study also provide an opportunity to make a comparison between keystroke data collected from various demographic groups.

C. Classification Algorithms

Systems based on the keyboard pattern are examined in detail. In the literature, there are two different approaches, statistical methods, and machine learning methods. Degree of Disorder, Mean, and Euclidean Distance methods are frequently preferred among statistical methods [8], while Random Forest, Decision Trees, Extra Trees Classifier, Bagging Classifier, and Gradient Boosting Classifier algorithms are commonly used in machine learning methods [7]. The algorithms used differ according to the accuracy, evaluation metrics, data type, amount, and collection method [6]–[8].

III. AUTHENTICATION SYSTEM BASED ON KEYSTROKE DYNAMICS

Among behavioral biometric systems, the accuracy of authentication methods using the keyboard pattern has considerably increased thanks to machine learning methods [9], [12]. It has been observed that problems can be solved quickly and more accurately with classical machine learning and artificial neural network-based deep learning methods than statistical methods. For this reason, studies in this field have increased, and it has been significant to design a successful model. This section details the steps of the keyboard pattern-based authentication process.

We design an authentication system based on keystroke dynamics and share the phases of our authentication system scheme in Figure 6. In this system, we develop our models using different machine learning algorithms to classify two different databases.

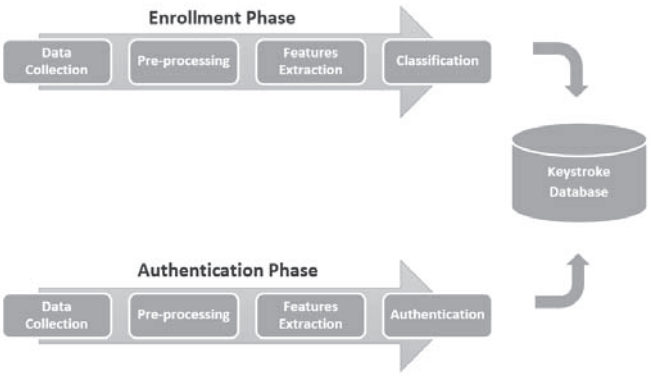

Fig. 2. Authentication system scheme.

A. Data Collection Application

In this study, two different databases were used to evaluate models. One of them is the database collected in the study presented in [11]. While creating this database, 51 different users entered the ".tie5Roanl" password 400 times, and the result of this study, 20,400 data were obtained. For second database, we developed a web application given in Figure 3 and collected keystroke data of a company employees. They were informed in detail about the content and usage area of the information collected. And we got permission from them for the use of their data. Also, in the first step of the data collection, we warned them not to use the passwords they use in any account for the security of their data.

Fig. 3. Data collection application.

In the register page defined in Figure 3, the user's information such as surname, gender, right or left-hand usage, age range, email address, and password was requested. While the users fill out the form, the keypress and release times of the users are recorded in the database, and then the user's data are anonymized. On the login page, keystroke data is collected while users log in to the system with their passwords. After logging in, we requested users to type the ".tie5Roanl" fixed password used in the benchmark database three times. While the users were typing the fixed password, the keyboard patterns were saved in the database. The reason for taking the password used in the benchmark database as a fixed password on the home page is to obtain results comparable with the studies in the previous studies.

Personal information such as name, surname, password, gender, right or left-hand usage, and age range entered in step 1 were collected for use in future studies [13]. In this study, the user's email address was taken as the primary key, and the keyboard patterns were recorded while entering the ".tie5Roanl" password. Finally, the classification process of the users was carried out through the keyboard patterns.

B. Data Preprocessing and Visualization

In order to increase the accuracy of the models, data preprocessing steps such as data quality evaluation, feature collection, feature sampling, dimension reduction were carried out. In this section, we gave information about our data preprocessing step and visualized the average keyboard patterns of the users. However, we have not shared details due to page constraints.

With the developed web application, 8819 login records were collected from 54 different users in 3 months. First, we deleted the rows with negative or NaN values and left 7905 records. During the data cleaning phase, we did not fill in the missing values with the average values of the columns as it negatively affects the user profile. We then calculated and visualized the average keyboard patterns of all users according to the Down-Down (DD), Up-Down (UD), and HOLD metrics.

The graphs defined in Figures 4, 5, and 6 illustrate the keystroke patterns created by taking the mean of the keystroke data for all users. Figure 4 shows the time elapsed between the user releasing a key and pressing the next key when typing the password ".tie5Roanl". This graphic shows the average values of the UD times of all users. Each line represents the average UD Time values of a user in the database.

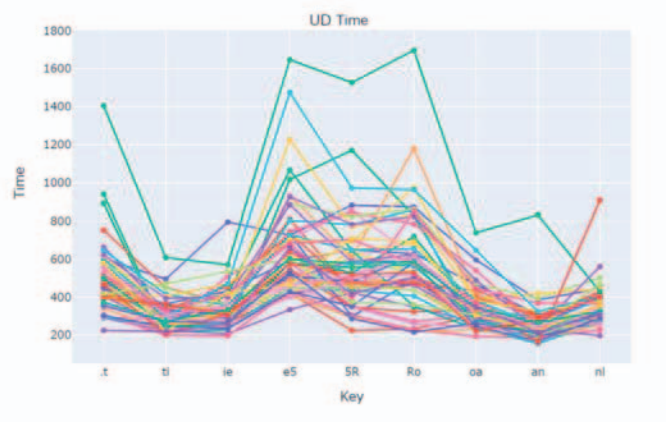

Fig. 4. An average keystroke pattern for each user in our database according to UD Time.

It is clearly seen that the keyboard patterns of different users given in Figures 4 and 5 are dissimilar from each other. These user data were passed through data pre-processing steps such as cleaning missing data, cleaning noises, and deviations in the data. First of all, among 7905 data collected from 54 people, people with less than 100 records were excluded. Among the users, the user with the least number of logins has 105 records, while the user with the highest number

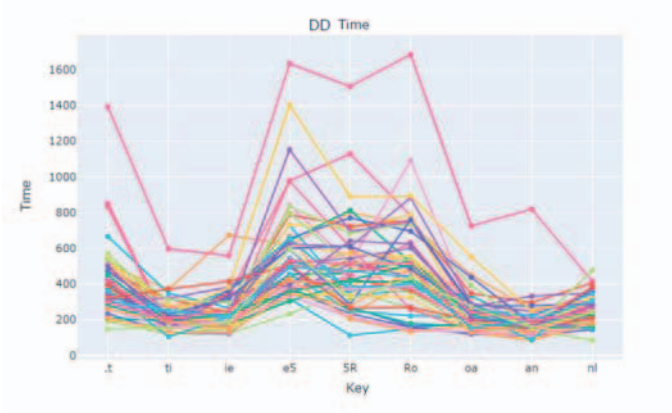

Fig. 5. An average keystroke pattern for each user in our database according to DD Time.

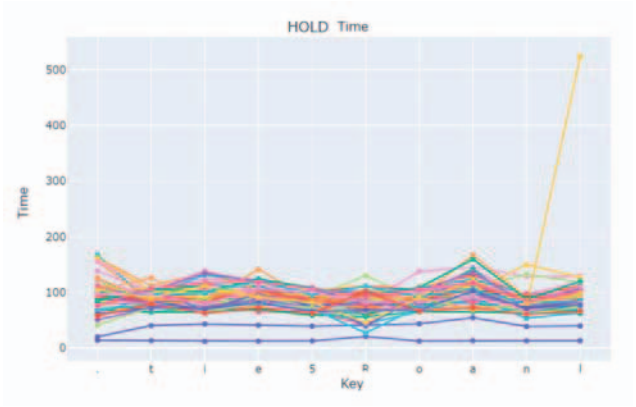

Fig. 6. An average keystroke pattern for each user in our database according to HOLD Time.

of logins has 273 records. And then, according to these graphics, maximum and minimum values for time metrics were defined. Subsequently, users deviating from the mean were removed from the database. At the end of the pre-processing phase, 35 regular users were selected among 54 users for the classification. Only two deviations were found in the average values of users who log in regularly and have more than 100 records.

C. The Classification Algorithms and Results

In this study, classification and regression-based basic models were developed using five classical machine learning algorithms (Random Forest, Decision Trees, SVM, and KNN), which were observed to have the highest accuracy in the previous studies.

While improving a machine learning model, we make a web application shown in Figure 3 for the data collection step. Then, the user was asked to enter the fixed ".tie5Roanl" password information three times in the application. During this process, we record the time the user presses the keys and calculated the user's transition times between keys. Finally, a behavioral pattern was obtained for a total of 54 users. After the data preprocessing step, 35 users were selected among 54 users.

In the classification step, the database was divided into two parts as train and test data. Kfold cross-validation was used while separating the data. Models were created using the most widely used machine learning algorithms in the literature and trained with a train dataset. Then, the test dataset was classified to evaluate the models. To evaluate the models, we selected three different time metrics as features: Hold Time, Down-Down Time, and Up-Down Time. In this study, two different methods were tried in the modeling phase. In the first method, eight different machine learning algorithms were used to classify keystroke data. Then, the super learner model, the ensemble model consisting of these eight machine learning algorithms, was used in the classification step. The accuracy scores of these machine learning algorithms and ensemble models are shown in Table I for both databases. To understand the effect of fold values in our database, we tried different k values and shared the results in Table I.

TABLE I
COMPARISON OF MODEL RESULTS FOR OUR AND BENCHMARK DATABASES

Model	Benchmark DB	Our DB		
	$k = 7$	$k = 5$	$k = 7$	$k = 10$
Super Learner	0.916	0.957	0.962	0.960
Bagging Classifier	0.841	0.937	0.940	0.946
Decision Tree Classifier	0.720	0.914	0.895	0.893
Random Forest Classifier	0.871	0.857	0.872	0.890
Extra Trees Classifier	0.851	0.837	0.845	0.840
Logistic Regression	0.741	0.718	0.687	0.699
SVC	0.210	0.613	0.600	0.577
K Neighbors Classifier	0.367	0.618	0.590	0.589
Gaussian NB	0.664	0.574	0.548	0.575

In Table I, we share possible k values. If the k value is too high, the accuracy score decreases because not enough data is distributed to the k fold. In cases where it is very small, the use of the k fold approach does not bring any benefit. Therefore, the most suitable value is obtained by trying different k values. In the second method, the Pycaret library was used for the classification phase. The results of the best machine learning models in this library are given in Table II for our database.

After that, the benchmark dataset presented in [11] was classified with our models. The results of this classification are shared in Table III.

According to the results of the classification process, it was observed that tree-based algorithms are the best model for both databases. Tree-based algorithms had an average accuracy of 94% and 93% in our database and benchmark dataset, respectively. It is predicted that there will be a rapid rise in this accuracy when the size of our dataset increases. Consequently, even if our database has less data than the benchmark database, it has higher accuracy than the benchmark database. It shows that developing the models is a critical phase and the amount of data collected is a considerable criterion for the performance of the model. For this reason, in the literature, it is recommended

TABLE II
PYCARET - BEST MODEL RESULTS FOR OUR DATABASE

Model	Accuracy	AUC	Recall	Prec.	F1
lgbm	0.969	0.999	0.968	0.971	0.969
gbc	0.950	0.999	0.946	0.953	0.950
rf	0.929	0.997	0.925	0.934	0.929
et	0.926	0.997	0.922	0.932	0.926
dt	0.894	0.945	0.890	0.900	0.893
lr	0.650	0.954	0.646	0.652	0.643
lda	0.638	0.959	0.625	0.653	0.629
knn	0.605	0.878	0.594	0.630	0.601
nb	0.604	0.940	0.609	0.655	0.600

Remark: lgbm = light gradient boosting machine, gbc = gradient boosting classifier, rf = random forest classifier, et = extra trees classifier, dt = decision tree classifier, lr = logistic regression, lda = linear discriminant analysis, knn = k neighbors classifier, nb = naive bayes

TABLE III
PYCARET - BEST MODEL RESULTS FOR BENCHMARK DATABASE

Model	Accuracy	AUC	Recall	Prec.	F1
et	0.938	0.997	0.938	0.941	0.938
lgbm	0.938	0.998	0.938	0.941	0.938
rf	0.930	0.997	0.930	0.932	0.929
gbc	0.915	0.998	0.915	0.919	0.915
lda	0.773	0.984	0.773	0.777	0.770
knn	0.737	0.928	0.737	0.758	0.737
qda	0.719	0.947	0.719	0.737	0.721
lr	0.712	0.950	0.711	0.716	0.701
dt	0.702	0.848	0.702	0.708	0.701
nb	0.675	0.957	0.675	0.696	0.670

Remark: et = extra trees classifier, lgbm = light gradient boosting machine, rf = random forest classifier, gbc = gradient boosting classifier, lda = linear discriminant analysis, knn = k neighbors classifier, qda = quadratic discriminant analysis lr = logistic regression, dt = decision tree classifier, nb = naive bayes

at least 20 up-to-date records should be kept for each user. When enough data is collected, keystroke dynamics between behavioral biometric systems reach the expected accuracy for two-factor authentication systems. For this reason, we foresee that its use will become more widespread in the future and will work in the background in many systems.

D. Limitations

We observe that there are inconsistencies in key press times due to the use of different devices in data collection and authentication steps [14]. Therefore, we examine the relationship between the accuracy of the algorithms and data collection techniques in the literature. As a solution to this, we try to increase the amount of data collected from various devices in the database. Then, the effect of external factors on the models was determined and was tried to be minimized.

In the data collection step, it is necessary to store and process personal data collected from users, taking into account legal restrictions. For this reason, our legal obligations were investigated in detail, and a report was prepared regarding this. Then, users were informed about the intended use of the collected data and the users' consent was obtained. Finally, before the data is processed, it has been anonymized within the scope of personal data protection law [15].

IV. RESULTS AND FUTURE STUDIES

It is believed that authentication systems based on keystroke dynamics may be preferred by public and private institutions that use systems where financial statements or sensitive personal information are kept, and identity control is significant. Nowadays, two-factor authentication systems are used in these sectors, and systems such as SMS authentication, which is more costly and less secure, are preferred as the second step. Therefore, we think that authentication systems based on keystroke dynamics will be a viable alternative for two-factor authentication.

As a result of this study, machine learning models were trained using the database we created and reached an average of 94% accuracy in the test phase. In future studies, we aim to develop fast and efficient models based on deep learning algorithms such as Long-Short Term Memory Networks (LSTM) based on Recurrent Neural Networks (RNN). Besides, we think that to establish new models considering the demographic characteristics collected from the users and to extract the relationships between features.

V. ACKNOWLEDGMENT

The authors are partially supported by TÜBITAK under grant no.7201256.

REFERENCES

[1] NIST, "Security and privacy controls for information systems and organizations," National Institue of Standarts and Technology, Gaithersburg, Maryland, USA, Tech. Rep. NISTIR 800-53 Revision 5, 2020.
[2] NIST, "Digital identity guidelines authentication and lifecycle management," National Institue of Standarts and Technology, Gaithersburg, Maryland, USA, Tech. Rep. NISTIR 800-63B, 2017.
[3] S. Boonkrong, *Multi-factor Authentication*. Berkeley, CA: Apress, 2021, pp. 133–162.
[4] BDDK, "Bankaların bilgi sistemleri ve elektronik bankacılık hizmetleri hakkında yönetmelik," Bankacılık Düzenleme ve Denetleme Kurumu, 2020.
[5] R. P. Jover, "Security analysis of sms as a second factor of authentication: The challenges of multifactor authentication based on sms, including cellular security deficiencies, ss7 exploits, and sim swapping," *Queue*, vol. 18, no. 4, p. 37–60, 2020.
[6] S. Dargan and M. Kumar, "A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities," *Expert Systems with Applications*, vol. 143, p. 113114, 2020.
[7] F. Fernández de Vega, D.-L. Yang, P. S. Teh, A. B. J. Teoh, and S. Yue, "A survey of keystroke dynamics biometrics," *The Scientific World Journal*, 2013.
[8] M. L. Ali, C. C. Tappert, M. Qiu, and J. V. Monaco, "Authentication and identification methods used in keystroke biometric systems," in *2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems*, 2015, pp. 1424–1429.
[9] M. L. Ali, J. V. Monaco, C. C. Tappert, and M. Qiu, "Keystroke biometric systems for user authentication," *Journal of Signal Processing Systems*, vol. 86, no. 2–3, p. 175–190, 2017.
[10] G. Pahuja and T. N. Nagabhushan, "Biometric authentication identification through behavioral biometrics: A survey," in *2015 International Conference on Cognitive Computing and Information Processing(CCIP)*, 2015, pp. 1–7.

[11] K. S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in *2009 IEEE/IFIP International Conference on Dependable Systems Networks*, 2009, pp. 125–134.
[12] I. Velásquez, A. Caro, and A. Rodríguez, "Authentication schemes and methods: A systematic literature review," *Information and Software Technology*, vol. 94, pp. 30 – 37, 2018.
[13] J. Solano, L. Camacho, A. Correa, C. Deiro, J. Vargas, and M. Ochoa, "Combining behavioral biometrics and session context analytics to enhance risk-based static authentication in web applications," *International Journal of Information Security*, 2020.
[14] N. Raul, R. Shankarmani, and P. Joshi, "A comprehensive review of keystroke dynamics-based authentication mechanism," in *International Conference on Innovative Computing and Communications*, A. Khanna, D. Gupta, S. Bhattacharyya, V. Snasel, J. Platos, and A. E. Hassanien, Eds. Singapore: Springer Singapore, 2020, pp. 149–162.
[15] KVKK, "Data protection in Turkey," Kişisel Verileri Koruma Kurumu, 2019.