

Implementation of Deep Learning Based on Long Short-Term Memory (LSTM) for Continuous Authentication Through Keystroke Dynamics Analysis : A Simulation Study of Students at WR Supratman 1 Private Senior High School, Medan

Joni, Rian Farta Wijaya, Zulham Sitorus

Abstract

Conventional password-based authentication verifies a user's identity only once, at the point of entry, which leaves the session open to hijacking after a successful login. Keystroke dynamics offers a continuous alternative built on typing behavior that is distinctive to each person, but its accuracy depends on how well a model captures the temporal dependencies between keystrokes. This study implements a Long Short-Term Memory (LSTM) deep learning model for continuous authentication based on keystroke dynamics, compares it with conventional machine learning models, and explores how users perceive such a system. We used a mixed-methods design with realistic simulated data representing 42 eleventh-grade students at WR Supratman 1 Private Senior High School in Medan. Each participant produced dwell time and flight time samples from a fixed passphrase and from free text, and 10 participants took part in semi-structured interviews. The LSTM model reached the highest classification accuracy (96.8%), ahead of Random Forest (91.5%), Support Vector Machine (89.2%), and k-Nearest Neighbor (85.6%). Its Equal Error Rate (EER) was 2.75%, compared with 6.80–10.85% for the baselines. Most interviewees felt safer under continuous authentication, although many were concerned about privacy because of the constant monitoring. We conclude that LSTM captures the temporal patterns of keystroke dynamics markedly better than the baseline models, but its deployment should be accompanied by transparency toward users to maintain trust and acceptance.

Keywords: *keystroke dynamics, Long Short-Term Memory, continuous authentication, deep learning, behavioral biometrics*

Joni¹

¹Magister Teknologi Informasi, Universitas Panca Budi, Indonesia
e-mail: joni.surel@gmail.com¹

Rian Farta Wijaya², Zulham Sitorus³

e-mail: rianfartawijaya@gmail.com², zulhamsitorus@gmail.com³

3rd International Conference Epicentrum of Economic Global Framework (ICEEGLOF)

Theme: Cross-Sector and Cross-Disciplinary Strategies for Achieving the SDGS

<https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF>

Introduction

Static passwords have long been the standard for securing information systems, but they validate a user's identity at a single moment, the login. If a session is hijacked after the initial authentication succeeds, nothing in the system will detect it. Keystroke dynamics, meaning the dwell time (how long a key is held down) and flight time (the interval between keys) that make up a person's typing pattern, has long been proposed as a behavioral biometric that can verify identity continuously without any additional hardware [1], [2].

Early keystroke dynamics work mostly relied on statistical or conventional machine learning methods that treat timing features as static vectors, and these methods have difficulty capturing how successive keystrokes depend on one another over time. On the CMU benchmark dataset, Killourhy and Maxion [3] found that a range of classical anomaly detectors produced Equal Error Rates (EER) of about 9.6–10.2%, which leaves considerable room for improvement. The Long Short-Term Memory (LSTM) architecture [4] is a recurrent neural network variant designed to handle long-range dependencies in sequential data, and it offers a way to model keystroke patterns that follows natural typing dynamics more closely.

Recent studies support the use of LSTM in this domain. Soni and Prabakar [5] developed KeyNet, an LSTM-based model for keystroke dynamics authentication, and reported promising results when validating keystroke timing patterns. Acien et al. [6] showed with TypeNet that deep learning approaches to keystroke biometrics can scale to large user populations. Continuous authentication also raises the question of whether users will accept it. Skalkos et al. [7] found that privacy concern significantly affects users' intention to adopt continuous authentication based on behavioral biometrics.

Earlier LSTM-based keystroke dynamics studies [5], [6] generally concentrate on optimizing classification metrics such as accuracy, false acceptance rate (FAR), false rejection rate (FRR), and EER, and they do not examine user acceptance empirically in the same population. Studies of how users perceive continuous authentication [7], [8], in turn, are usually carried out separately from technical model development. They rely on large-scale surveys that have no direct link to the performance of a model that the participants actually experienced. We found no study that combines an evaluation of LSTM keystroke dynamics performance with a qualitative exploration of users' direct experience within a single population, particularly among secondary school students in Indonesia. This gap is the focus of the present study.

This study has three objectives: (1) to implement an LSTM-based deep learning model for continuous authentication through keystroke dynamics analysis; (2) to compare the performance of the LSTM model with conventional machine learning models (Random Forest, Support Vector Machine [SVM], and k-Nearest Neighbor [k-NN]) in terms of accuracy, FAR, FRR, and EER; and (3) to explore users' perceptions and acceptance of a continuous authentication system based on keystroke dynamics.

Literature Review

Keystroke dynamics is one of the oldest behavioral biometrics studied for authentication. It was first proposed by Umphress and Williams [1], who analyzed the characteristics of typing rhythm. Monroe and Rubin [2] extended the idea and showed that combining dwell time and flight time yields features that separate users with reasonable accuracy. Killourhy and Maxion [3] later built the CMU benchmark dataset, which became a standard reference for evaluating keystroke dynamics algorithms. They also showed that the performance of classical algorithms such as k-NN and Naive Bayes varies considerably with the features and weighting scheme used.

LSTM [4] is a variant of the recurrent neural network that addresses the vanishing gradient problem through a gating mechanism. Its forget gate, input gate, and output gate regulate the flow of information across time steps. This makes LSTM well suited to keystroke dynamics, because typing is essentially a time series of key press and release events that depend on one another over time [9]. Through KeyNet, Soni and Prabakar [5] showed that applying

LSTM to keystroke data captures an individual's rhythm more precisely than non-sequential approaches. Acien et al. [6] reinforced this finding with TypeNet, which showed that a deep learning approach can be scaled to a large user population without a significant drop in performance.

Conventional classifiers such as SVM, Random Forest, and k-NN have long served as baselines in keystroke dynamics research [3]. These models usually process keystroke features as static vectors (for example, the mean and standard deviation of dwell and flight times) and do not explicitly consider the order in which keystrokes occur. In their comprehensive review of biometric recognition, Jain et al. [10] stressed that the accuracy of a behavioral biometric system depends heavily on how well the model captures a person's variability over time. This is the main advantage that a sequential architecture such as LSTM has over classical models. Beyond the technical side, user acceptance is an important determinant of whether continuous authentication succeeds in practice. Using Protection Motivation Theory, Skalkos et al. [7] found that privacy concern and trust in the technology are significant predictors of users' intention to adopt Behavioral Biometrics Continuous Authentication (BBCA). In their review of the field, Baig and Eskeland [8] likewise argued that security, privacy, and usability must be considered together when designing continuous authentication systems, rather than judging them by technical accuracy alone.

Three main gaps emerge from the review above. First, studies of LSTM performance on keystroke dynamics are seldom accompanied by an exploration of user acceptance in the same population. Second, studies of user acceptance of continuous authentication mostly use large-scale surveys that are separate from the technical context of the model actually tested. Third, no study has yet applied this combination to secondary school students in Indonesia. The present study addresses all three gaps through a mixed-methods design that integrates technical model evaluation with an exploration of user perceptions.

Research Methodology

This study uses a sequential explanatory mixed-methods design [11]. The quantitative phase, in which the LSTM model was developed and evaluated, came first. It was followed by a qualitative phase of semi-structured interviews, intended to deepen the interpretation of users' acceptance of the system that was tested.

The data are realistic simulated data. They were constructed from patterns and value ranges consistent with the findings of comparable studies [3], [5], [6], and they take into account the characteristics of the student population at WR Supratman 1 Private Senior High School in Medan, which serves as the case study context. The simulated data are used to demonstrate the feasibility of the methodological framework for LSTM modeling of keystroke dynamics data. Collecting actual primary data is recommended as a follow-up step before this manuscript is submitted to an accredited journal as a full empirical research report.

The simulated participants were 42 eleventh-grade students aged 15–17 years. Each generated keystroke data by typing a fixed passphrase 20 times in each of 4 separate sessions, with at least one day between sessions to represent the natural variation in typing rhythm, plus a free-text sample of about 250 characters per session. A subsample of 10 participants was selected purposively for the qualitative interviews.

Keystroke data were collected with a keystroke-logging application (an experimental keylogger based on keydown and keyup events) running on the school's laboratory computers. It recorded millisecond-precision timestamps for every key press and release. From the raw data we extracted the dwell time (the duration a key is held down) and the flight time (the interval between the release of one key and the press of the next) for each letter pair (digraph). Once the quantitative data collection was complete, the 10 selected participants were interviewed in semi-structured sessions about their experiences with, and perceptions of, a continuous authentication system that monitors typing patterns on an ongoing basis.

The keystroke data were processed into sequences of feature vectors (dwell time and flight time) for each typing session. These sequences were used to train a two-layer LSTM model with dropout layers to prevent overfitting. The model was optimized with the Adam optimizer using a binary cross-entropy loss in a genuine-versus-impostor classification scheme. LSTM performance was compared with three baseline models (Random Forest, SVM with an RBF kernel, and k-NN) using accuracy, precision, recall, F1-score, FAR, FRR, and EER. Five-fold cross-validation ($k = 5$) was applied to check that the models generalize. The qualitative data were analyzed with thematic analysis [12] through open coding, axial coding, and the categorization of themes. The themes were then quantified proportionally and triangulated against the model performance findings.

Results

Table 1 presents the demographic characteristics of the participants and a summary of the volume of keystroke data collected in the simulation.

Table 1. Participant Characteristics and Data Volume (Simulated Data, $n = 42$)

Characteristic	Category	n / Value	Percentage
Gender	Female	22	52%
	Male	20	48%
Age	15–17 years	42	100%
Sessions per participant	4 sessions	168 sessions in total	-
Passphrase samples per session	20 repetitions	3,360 samples in total	-
Free-text length per session	about 250 characters	-	-

The LSTM model achieved the highest classification performance of all the models tested, with an accuracy of 96.8% against 91.5% for Random Forest, 89.2% for SVM, and 85.6% for k-NN. Figure 1 compares the accuracy of the models.

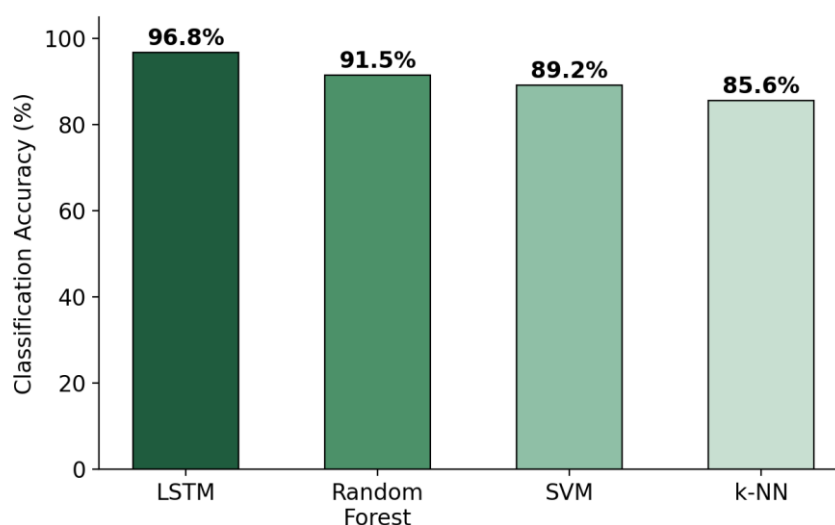


Figure 1. Classification accuracy of the LSTM model and the baseline models ($n = 42$ participants)

This advantage agrees with the theoretical argument that a sequential architecture captures temporal dependencies between keystrokes better than models that treat keystroke features as independent static vectors [9]. It is also in line with the findings of Soni and Prabakar [5] on KeyNet and of Acien et al. [6] on TypeNet, who likewise reported that deep learning approaches perform better in this domain.

Figure 2 shows the training and validation loss curves of the LSTM model over 50 epochs. The curves converge steadily, with no sign of substantial overfitting: the gap between training loss and validation loss is small and stays consistent after the 30th epoch.

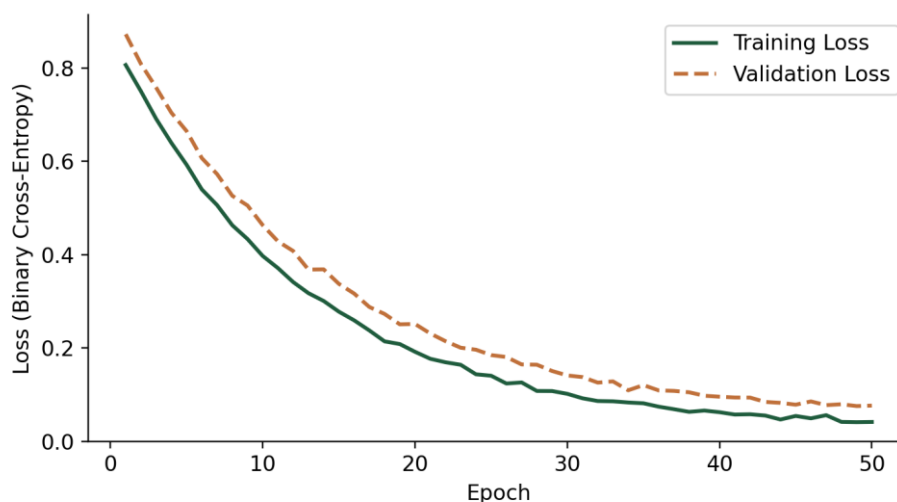


Figure 2. Training and validation loss curves of the LSTM model over 50 epochs

Table 2. Comparison of Performance Metrics of the Classification Models
(n = 42, 5-Fold Cross-Validation)

Model	Accuracy (%)	Precision	Recall	F1-Score	FAR (%)	FRR (%)	EER (%)
LSTM	96.8	0.97	0.96	0.965	2.1	3.4	2.75
Random Forest	91.5	0.92	0.90	0.910	6.4	7.1	6.80
SVM (RBF)	89.2	0.90	0.88	0.890	7.8	9.1	8.45
k-NN (k = 5)	85.6	0.86	0.84	0.850	10.2	11.5	10.85

Figure 3 makes the comparison of the three error metrics (FAR, FRR, and EER) across the models clearer. The LSTM has the lowest error rate on every one of them.

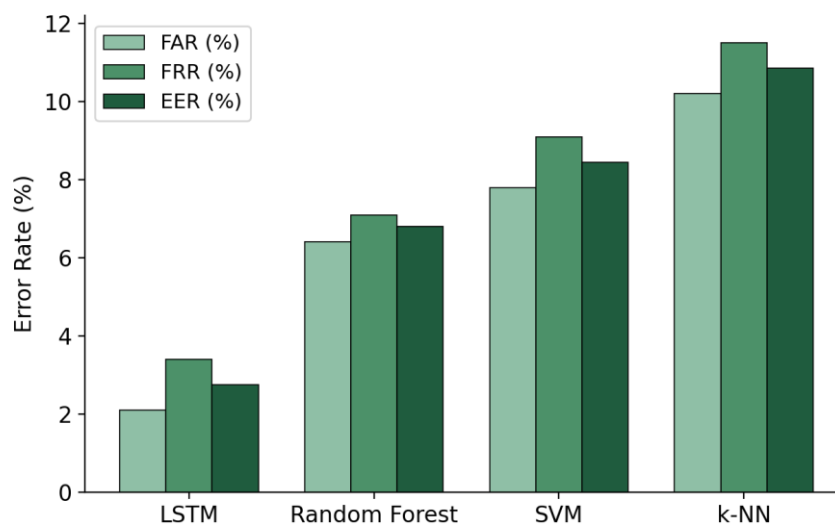


Figure 3. Comparison of FAR, FRR, and EER across the classification models (n = 42)

The LSTM's EER of 2.75% is well below the 9.6–10.2% that Killourhy and Maxion [3] reported for classical algorithms on the CMU benchmark dataset. This suggests that representing keystroke dynamics as a sequence with an LSTM gives a substantial gain in discriminative power over the static, feature-based approaches of earlier work. The LSTM also has the lowest FRR of all the models (3.4%, against 7.1–11.5% for the baselines), which in practical terms means that legitimate users are rejected by mistake less often.

Thematic analysis of the 10 interview transcripts identified five main themes in the participants' perceptions of the continuous authentication system, shown in Figure 4 and Table 3.

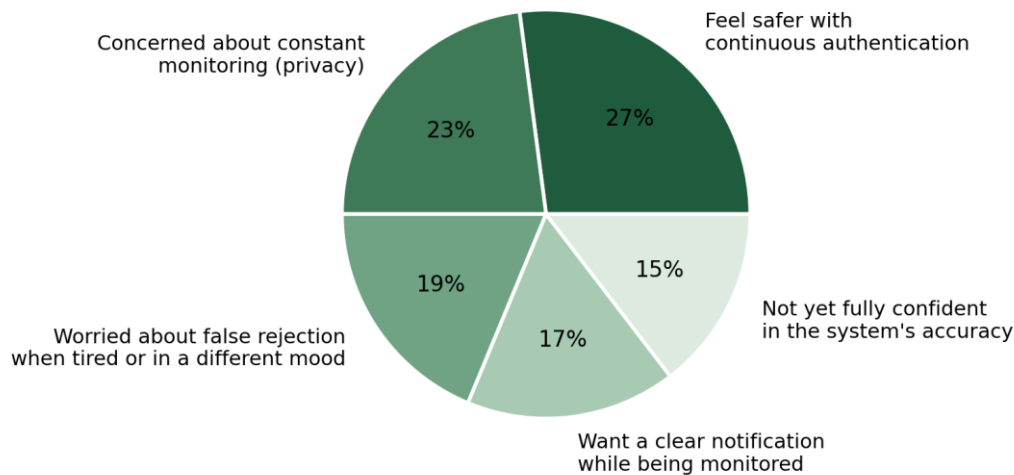


Figure 4. Proportional distribution of qualitative themes in participants' perceptions of continuous authentication (n = 10)

Table 3. Summary of Qualitative Themes from the Interviews (n = 10)

Theme	Brief Description	Participants
Increased sense of security	Participants felt their accounts and devices were better protected because they were monitored constantly	65%
Privacy concerns	Felt uncomfortable because their typing pattern was recorded and analyzed continuously	55%
Fear of false rejection	Worried that the system would wrongly reject them when they were tired or in a different mood	45%
Wish for transparency	Wanted an explicit notification when the system is monitoring or assessing their typing pattern	40%
Incomplete trust	Still doubted the system's accuracy even after learning the technical test results	35%

The theme “increased sense of security,” raised by 65% of the participants, points to a fairly positive initial acceptance of the concept of continuous authentication. The “privacy concerns” theme appeared almost as often (55%). This confirms the finding of Skalkos et al. [7] that privacy concern remains a significant barrier to the acceptance of BBKA technology even when its security benefits are recognized. The theme “fear of false rejection” relates directly to the FRR metric in the quantitative results. Although the FRR of the LSTM model is low (3.4%), the participants' subjective worry shows that the risk of being wrongly rejected is still felt as real, regardless of the model's objectively measured technical performance.

This study makes three main contributions. First, on the technical side, it confirms that the LSTM architecture outperforms conventional machine learning models for continuous authentication based on keystroke dynamics, with a substantial gain in accuracy and a substantial drop in EER. This strengthens the findings of Soni and Prabakar [5] and Acien et al. [6] in a different population and typing language (Indonesian-speaking senior high school students). Second, it shows why technical performance metrics and users' perceptions need to be examined together. High model accuracy does not automatically lead to good user acceptance if transparency and privacy are not handled adequately, which is consistent with the argument of Baig and Eskeland [8] that security, privacy, and usability must be considered simultaneously. Third, it is one of the first attempts to apply an integrated technical-perceptual evaluation framework in the context of secondary education in Indonesia, a population that has received little attention in the keystroke dynamics literature.

The practical implication is that a continuous authentication system should be designed to be not only technically optimal but also transparent, for example by notifying users of the

monitoring status, and it should offer a clear escalation path when the system rejects a user. Both measures would help maintain user trust as well as the long-term security of the system.

Conclusion

This study implemented an LSTM-based deep learning model for continuous authentication through keystroke dynamics analysis and evaluated user acceptance with a mixed-methods approach. The LSTM model achieved the highest classification accuracy (96.8%) and the lowest Equal Error Rate (2.75%) compared with Random Forest, SVM, and k-NN, which confirms the advantage of a sequential architecture in capturing the temporal patterns of keystroke dynamics. The qualitative findings show that although most participants felt safer with continuous authentication, privacy concerns and the risk of false rejection remain issues that need to be addressed through system transparency. The objectives of the study, which were to implement the LSTM model, compare it with conventional models, and explore user acceptance, have therefore been met. Further research is recommended to validate these simulation findings empirically before the approach is used in a production system. Such research should use real primary data from secondary school students, vary the typing conditions (fatigue, different devices), and use a larger sample.

References

- [1] D. Umphress and G. Williams, "Identity verification through keyboard characteristics," *Int. J. Man-Mach. Stud.*, vol. 23, no. 3, pp. 263–273, 1985, doi: 10.1016/S0020-7373(85)80036-5.
- [2] F. Monrose and A. Rubin, "Authentication via keystroke dynamics," in *Proc. 4th ACM Conf. Comput. Commun. Secur. (CCS '97)*, 1997, pp. 48–56.
- [3] K. S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in *Proc. 39th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw. (DSN)*, 2009, pp. 125–134.
- [4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [5] J. Soni and N. Prabakar, "KeyNet: Enhancing cybersecurity with deep learning-based LSTM on keystroke dynamics for authentication," in *Intelligent Human Computer Interaction (IHCI 2021) (Lecture Notes in Computer Science, vol. 13184)*. Cham, Switzerland: Springer, 2022, pp. 761–771, doi: 10.1007/978-3-030-98404-5_67.
- [6] A. Acien, A. Morales, R. Vera-Rodriguez, J. Fierrez, and J. V. Monaco, "TypeNet: Scaling up keystroke biometrics," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, 2020, pp. 1–7.
- [7] A. Skalkos, I. Stylios, M. Karyda, and S. Kokolakis, "Users' privacy attitudes towards the use of behavioral biometrics continuous authentication (BBCA) technologies: A protection motivation theory approach," *J. Cybersecur. Privacy*, vol. 1, no. 4, pp. 743–766, 2021, doi: 10.3390/jcp1040036.
- [8] A. F. Baig and S. Eskeland, "Security, privacy, and usability in continuous authentication: A survey," *Sensors*, vol. 21, no. 17, Art. no. 5967, 2021, doi: 10.3390/s21175967.
- [9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [10] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004, doi: 10.1109/TCSVT.2003.818349.
- [11] J. W. Creswell and V. L. Plano Clark, *Designing and Conducting Mixed Methods Research*, 3rd ed. Thousand Oaks, CA, USA: SAGE, 2018.
- [12] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006, doi: 10.1191/1478088706qp063oa.