Vilniaus universitetas
Duomenų mokslo ir skaitmeninių technologijų institutas

Informatikos krypties doktorantų ataskaitinė konferencija
Veiklos ataskaita už 2023 m. rugsėjo 27 d. – 2024 m. rugsėjo 30 d.
II pusmetis

Giliuoju mokymusi pagrįstas klavišų paspaudimų dinamikos autentifikavimas vidinių grėsmių aptikimui ypatingos svarbos infrastruktūroje

dokt. Arnoldas BUDŽYS – Informatika N 009
Studijų metai: IV
Darbo vadovas: dr. Viktor Medvedev
Doktorantūros pradžios ir pabaigos metai: 2020–2024

2023–2024 m.

Studijų planas ir jo vykdymo suvestinė

Studijų metai	Egzaminai	
	Planas	Įvykdyta
I (2020/2021)	1	1
II (2021/2022)	3	3
III (2022/2023)		
IV (2023/2024)		
Iš viso:	4	4

Studijų metai	Dalyvavimas konferencijose				Publikacijos					
	Tarptautinėse		Nacionalinėse		Su citavimo rodikliu			Be citavimo rodiklio		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė	Planas	Įvykdyta	Būklė
I (2020/2021)										
II (2021/2022)	1	1	1	1				1	0	
III (2022/2023)	1	1+1**	0	1	1	1	Publikuota	1	2	Publikuota
IV (2023/2024)			**0**	**1**	**1**	**2**	**Publikuota**			
Iš viso:	2	2+1**	1	3	2	1		1	2	

*HCI2023 – publikuota (CA WoS, Springer), CISTI2023 – publikuota (CA WoS, IEEE).

** Prisidėta prie pranešimo bei straipsnio konferencijų medžiagoje parengimo.

Ataskaitinis studijų pusmetis
(IV: 2023/2024 – II pusmetis)

Publikacijos 2023/2024 (II pusmetis)			
Planas	**Įvykdyta**	**Būklė**	**Publikacijos tipas**
Artificial Intelligence Review (Springer Nature)	**Budžys, A.;** Kurasova, O.; Medvedev, V. Deep learning-based authentication for insider threat detection in critical infrastructure // Artificial intelligence review. Dordrecht : Springer Nature B.V. ISSN 0269-2821. eISSN 1573-7462. 2024, vol. 57, iss. 10, art. no. 272, p. [1-35]. DOI: 10.1007/s10462-024-10893-1.	Publikuota	CA WoS duomenų bazėje, turi cituojamumo rodiklį (**IF – 10.7**); Computer Science, Artificial Intelligence in SCIE edition – Q1
Computer Standards & Interfaces (Elsevier)	**Budžys, A.;** Kurasova, O.; Medvedev, V. Integrating deep learning and data fusion for advanced keystroke dynamics authentication // Computer standards & interfaces. Amsterdam : Elsevier B.V. ISSN 0920-5489. 2025, vol. 92, art. no. 103931, p. 1-14. DOI: 10.1016/j.csi.2024.103931.	Publikuota	CA WoS duomenų bazėje, turi cituojamumo rodiklį (**IF– 4.1**); Computer Science, Hardware & Architecture in SCIE – Q1; Computer Science, Software Engineering in SCIE edition – Q1

Doktorantūros studijų pasiekimai

Dalyvavimas tarptautinėse konferencijose	
1.	**Budžys, A.,** Kurasova, O., and Medvedev, V., „Deep learning-based prevention of insider threats using user behavioral keystroke biometrics“, 32nd European Conference on Operational Research (EURO XXXII)], Espoo, Finland, July 3-6, 2022.
2.	**Budžys, A.,** Kurasova, O., and Medvedev, V., “Behavioral Biometrics Authentication Using Siamese Neural Networks”, in HCI for Cybersecurity, Privacy and Trust 5th International Conference, HCI-CPT 2023, Held as Part of the 25th HCI International Conference, HCI2023, 2023, pp. 1–14.
3.*	**Medvedev, V.,** Budžys, A., and Kurasova, O. “Enhancing keystroke biometric authentication using deep learning techniques”, 2023 18th Iberian Conference on Information Systems and Technologies (CISTI), 20-23 June, Aveiro, Portugal, 2023 : proceedings. New York: IEEE, 2023.

*Prisidėta prie pranešimo bei straipsnio konferencijų medžiagoje parengimo.

Doktorantūros studijų pasiekimai

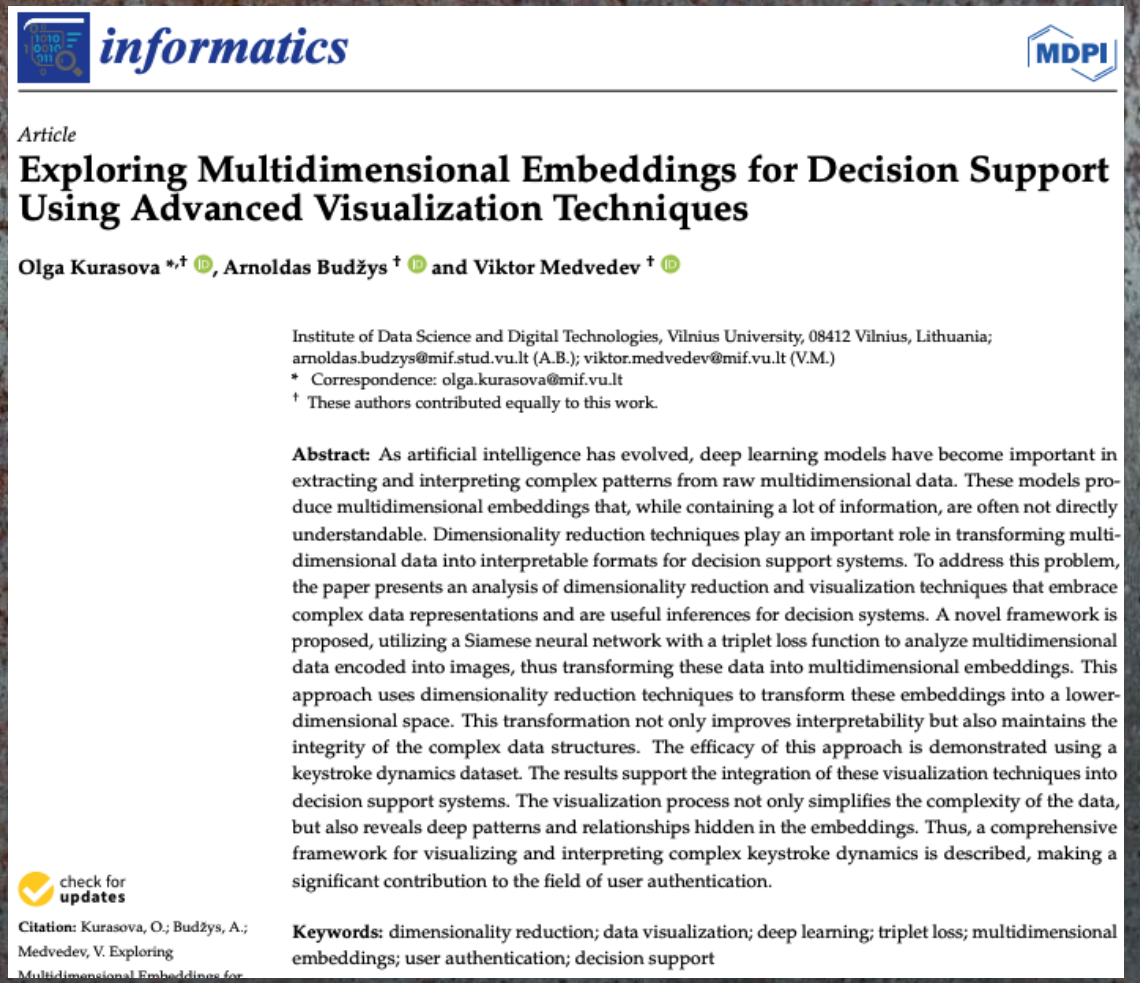
INFORMATICS-BASEL

Publisher name: MDPI

Journal Impact Factor™

3.4	3.1
2023	Five Year

JCR Category	Category Rank	Category Quartile
COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS *in ESCI edition*	61/169	Q2



informatics

MDPI

Article

Exploring Multidimensional Embeddings for Decision Support Using Advanced Visualization Techniques

Olga Kurasova *,† , Arnoldas Budžys † and Viktor Medvedev †

Institute of Data Science and Digital Technologies, Vilnius University, 08412 Vilnius, Lithuania; arnoldas.budzys@mif.stud.vu.lt (A.B.); viktor.medvedev@mif.vu.lt (V.M.)
* Correspondence: olga.kurasova@mif.vu.lt
† These authors contributed equally to this work.

Abstract: As artificial intelligence has evolved, deep learning models have become important in extracting and interpreting complex patterns from raw multidimensional data. These models produce multidimensional embeddings that, while containing a lot of information, are often not directly understandable. Dimensionality reduction techniques play an important role in transforming multidimensional data into interpretable formats for decision support systems. To address this problem, the paper presents an analysis of dimensionality reduction and visualization techniques that embrace complex data representations and are useful inferences for decision systems. A novel framework is proposed, utilizing a Siamese neural network with a triplet loss function to analyze multidimensional data encoded into images, thus transforming these data into multidimensional embeddings. This approach uses dimensionality reduction techniques to transform these embeddings into a lower-dimensional space. This transformation not only improves interpretability but also maintains the integrity of the complex data structures. The efficacy of this approach is demonstrated using a keystroke dynamics dataset. The results support the integration of these visualization techniques into decision support systems. The visualization process not only simplifies the complexity of the data, but also reveals deep patterns and relationships hidden in the embeddings. Thus, a comprehensive framework for visualizing and interpreting complex keystroke dynamics is described, making a significant contribution to the field of user authentication.

check for updates

Citation: Kurasova, O.; Budžys, A.; Medvedev, V. Exploring Multidimensional Embeddings for

Keywords: dimensionality reduction; data visualization; deep learning; triplet loss; multidimensional embeddings; user authentication; decision support

Publikacijos (tik su citavimo rodikliu)		
	Bibliografinis aprašas	Būklė
Emerging Sources Citation Index (ESCI) duomenų bazėje	Kurasova, O.; **Budžys, A.**; Medvedev, V. Exploring multidimensional embeddings for decision support using advanced visualization techniques // Informatics. Basel : MDPI. eISSN 2227-9709. 2024, vol. 11, iss. 1, art. no. 11, p. [1-17]. DOI: 10.3390/informatics11010011.	Publikuota

Doktorantūros studijų pasiekimai

ARTIFICIAL INTELLIGENCE REVIEW

Publisher name: SPRINGER

Journal Impact Factor ™

10.7	11.7
2023	Five Year

JCR Category	Category Rank	Category Quartile
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE *in SCIE edition*	9/197	Q1

Publikacijos (tik su citavimo rodikliu)		
	Bibliografinis aprašas	Būklė
Science Citation Index-Expanded (SCIE) duomenų bazėje	**Budžys, A.,** Kurasova, O., Medvedev, V. Deep Learning-Based Authentication for Insider Threat Detection in Critical Infrastructure. Artificial Intelligence Review (2024). DOI: 10.1007/s10462-024-10893-1. [Science Citation Index Expanded (SCIE) (Web of Science), IF – 10,7; CA WoS duomenų bazėje; Computer Science, Artificial Intelligence in SCIE edition – Q1	Publikuota

Doktorantūros studijų pasiekimai

COMPUTER STANDARDS & INTERFACES

Publisher name: ELSEVIER

Journal Impact Factor™

4.1	3.2
2023	Five Year

JCR Category	Category Rank	Category Quartile
COMPUTER SCIENCE, HARDWARE & ARCHITECTURE *in SCIE edition*	10/59	Q1
COMPUTER SCIENCE, SOFTWARE ENGINEERING *in SCIE edition*	20/131	Q1

Computer Standards & Interfaces 92 (2025) 103931

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

Integrating deep learning and data fusion for advanced keystroke dynamics authentication

Arnoldas Budžys *, Olga Kurasova, Viktor Medvedev

Institute of Data Science and Digital Technologies, Vilnius University, Akademijos str. 4, Vilnius 08412, Lithuania

ARTICLE INFO

Keywords:
Cybersecurity
Keystroke dynamics
User authentication
Siamese neural network
Data fusion
Critical infrastructure

ABSTRACT

By enhancing user authentication protocols, especially in critical infrastructures vulnerable to complex cyberthreats, we present an advanced approach that integrates a deep learning-based model and data fusion techniques applied to analyze keystroke dynamics. With the growing need for robust security measures, especially in critical infrastructure environments, traditional authentication mechanisms often fail to cope with advanced threats. Our approach focuses on the unique behavioral biometric characteristics of keystrokes, which offers promising opportunities to improve user authentication processes. We have developed a data fusion-based methodology that utilizes the unique features of keystroke dynamics combined with deep learning techniques to improve user authentication systems. Using the capabilities of data fusion and deep learning, the proposed methodology not only captures the complex behavioral biometrics inherent in keystroke dynamics but also addresses the challenges posed by varying password lengths and typing styles. We conducted extensive experiments on several fixed-text datasets, including the Carnegie Mellon University dataset, the KeyRecs dataset, and the GREYC-NISLAB dataset, with a total of approximately 54,000 password records. Comprehensive experiments on various datasets with different password lengths have shown that our approach is scalable and accurate for user authentication, which significantly improves the security of critical infrastructure. By using interpolation-based data fusion techniques to standardize the keystroke data to a uniform length and employing a Siamese neural network with a triplet loss function, the best equal error rate of 0.13281 was achieved for the unseen fused data. The integration of deep learning and data fusion effectively generalizes to different user profiles, demonstrating its adaptability and accuracy in authenticating users in different scenarios. The findings are crucial for improving security in sensitive applications, ranging from accessing personal devices to protecting critical infrastructure.

Publikacijos (tik su citavimo rodikliu)		
	Bibliografinis aprašas	Būklė
Science Citation Index-Expanded (SCIE) duomenų bazėje	**Budžys, A.,** Kurasova, O., Medvedev, V. Integrating Deep Learning and Data Fusion Based Authentication // Computer Standards & Interfaces . Elsevier B. V. ISSN 0920-5489. 2024, vol. 92, p. [1-14]. DOI: 10.1016/j.csi.2024.103931.	Publikuota

Doktorantūros mokslinių tyrimų ir disertacijos rengimo etapai

Darbo pavadinimas	Atlikimo terminai	Pastabos
Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):	2020 m. spalio mėn. – 2021 m. rugsėjo mėn.	Parengta
Mokslinio tyrimo vykdymas:		
2.1. Tyrimo metodikos sudarymas: 2.1.1. Tyrimo metodikos iškeltiems uždaviniams spręsti parinkimas; 2.1.2. Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką.	2021 m. spalio mėn. – 2022 m. sausio mėn.	Parengta
2.2. **Teorinis tyrimas:** 2.2.1. Mašininio mokymosi metodų, naudojamų kompiuterių tinkluose įsilaužimų prevencijai, tyrimas.	2022 m. vasario mėn. – 2022 m. rugsėjo mėn.	Parengta
2.3. **Empirinis tyrimas:** 2.3.1. Sudarytų metodų pritaikymas praktinių uždavinių sprendimui. 2.3.2. Gautų duomenų analizė, rezultatų apibendrinimas, išvadų parengimas.	2022 m. spalio mėn. – 2023 m. rugsėjo mėn.	Parengta

Doktorantūros mokslinių tyrimų ir disertacijos rengimo etapai

Darbo pavadinimas	Atlikimo terminai	Pastabos
Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas: 3.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas; 3.2. Analitinės disertacijos dalies parengimas; 3.3. Teorinės disertacijos dalies parengimas; 3.4. Eksperimentinės disertacijos dalies parengimas; 3.5. Bendrųjų išvadų formulavimas.	2023 m. spalio mėn. – 2024 m. gegužės mėn.	Parengtas juodraštis
Daktaro disertacijos parengimas ir svarstymas padalinyje	2024 m. – 2025 m.	
Daktaro disertacijos gynimas	2025 m.	

Disertacijos tema, tyrimo objektai ir tikslas

Disertacijos tema:

- Giliuoju mokymusi pagrįstas klavišų paspaudimų dinamikos autentifikavimas vidinių grėsmių aptikimui ypatingos svarbos infrastruktūroje.

Tikslas:

- Pasiūlyti ir įvertinti giliuoju mokymusi ir klavišų paspaudimų dinamika pagrįstą metodiką naudotojų autentifikavimui, siekiant pagerinti vidinių grėsmių aptikimą kritinės infrastruktūros sistemose.

Tyrimo objektai:

- Vartotojo sugeneruoti klaviatūros biometriniai skaitiniai duomenys, metodai vidinių grėsmių identifikavimui ir neteisėtų veiksmų užkardymui gerinant bendrą kibernetinį saugumą.
- Giliojo mokymosi metodų kūrimas ir taikymas naudotojų autentifikavimui pagal klavišų paspaudimų dinamiką kritinės infrastruktūros sistemose, panaudojant Siamo neuroninius tinklus su trigubų nuostolių funkcija.
- Klavišų paspaudimų duomenų transformavimas į vaizdines reprezentacijas metodai.

Tyrimo uždaviniai

- Atlikti išsamią analitinę literatūros apžvalgą, siekiant ištirti ir nustatyti veiksmingus naudotojų autentifikavimo metodus ypatingos svarbos infrastruktūros sistemose, daugiausia dėmesio skiriant elgsenos biometrikos, pavyzdžiui, klavišų paspaudimų dinamikos, naudojimui ypatingos svarbos infrastruktūros saugumui didinti.
- Ištirti klavišų paspaudimų dinamikos duomenų struktūrą ir charakteristikas, siekiant nustatyti metodus, kurie gali pagerinti naudotojo identifikavimą ir grėsmių aptikimą pagal naudotojo spausdinimo elgseną.
- Išnagrinėti skirtingus mašininio mokymosi metodus ir modelius, siekiant nustatyti jų gebėjimą aptikti vidines grėsmes pagal naudotojo spausdinimo elgseną ir įvertinti jų galimybes pagerinti autentifikavimo tikslumą.
- Sukurti naudotojo autentifikavimo kompleksinę sistemą ypatingos svarbos infrastruktūros padaliniuose, pagrįstą klavišų paspaudimų dinamika.
- Identifikuoti tinkamas tikslumo metrikas vartotojų autentifikavimo rezultatams įvertinti.
- Atlikti sukurtos kompleksinės sistemos efektyvumo įvertinimą, panaudojant viešai prieinamus klavišų paspaudimų dinamikos duomenų rinkinius.

Ginamieji teiginiai

- Klavišų paspaudimo dinamika, kaip elgsenos biometrijos forma, yra saugesnis ir adaptyvesnis naudotojo autentiškumo patvirtinimo būdas, palyginti su tradicinėmis slaptažodžiais grindžiamomis sistemomis. Unikalūs asmenų rašymo būdai gali būti veiksmingai naudojami siekiant aptikti vidines grėsmes ir užkirsti kelią neteisėtai prieigai prie ypatingos svarbos infrastruktūros sistemų.
- Klavišų paspaudimų dinamikos duomenų transformavimas į vaizdus pagerina gilaus mokymosi modelių efektyvumą. Ši transformacija leidžia tiksliau išgauti požymius, todėl sistema tiksliau atskiria teisėtus naudotojus nuo potencialių įsilaužėlių.
- Standartizavus skirtingus klavišų paspaudimų duomenis ir pritaikius duomenų transformavimo metodus, pavyzdžiui pasiūlytą GAFMAT, pagerėja sukurtos kompleksinės sistemos efektyvumas. Sukurti naudotojo autentifikavimo kompleksinę sistemą ypatingos svarbos infrastruktūros padaliniuose, pagrįstą klavišų paspaudimų dinamika.
- Pasiūlyta naudotojo autentifikavimo kompleksinė sistema praktiškai pritaikoma realioje kritinės infrastruktūros aplinkoje. Ji buvo įvertinta naudojant viešai prieinamus duomenų rinkinius ir eksperimentinius tyrimus, kurie įrodo jos pritaikomumą, universalumą ir aukšto lygio saugumo potencialą aplinkoje, kurioje naudotojų elgsena laikui bėgant gali kisti.

Mokslinis darbo naujumas

- Naujos klavišų paspaudimo duomenų transformavimo metodikos sukūrimas. *Gabor Filter Matrix Transformation (GAFMAT)* – skaitinių duomenų transformacija į vaizdinius formatus kas leidžia padidinti konvoliucinių neuroninių tinklų gebėjimą išgauti požymius iš transformuotų duomenų, o tai leidžia pagerinti autentifikavimo tikslumą.
- Pasiūlyta nauja skirtingų klavišų paspaudimų duomenų rinkinių standartizavimo metodika, skirtą klavišų paspaudimo požymiams normalizuoti įvairiuose duomenų rinkiniuose.
- Pasiūlyta kompleksinė sistema vidinių grėsmių aptikimui kritinės infrastruktūros padaliniuose, pritaikant giliuosius neuroninius tinklus elgsenos biometriniams duomenims analizuoti. Į sukurtą kompleksinę sistemą buvo integruotas duomenų transformavimas į vaizdų metodas GAFMAT, kuris leido pagerinti naudotojo autentifikavimo tikslumą.

Apibendrinimas ir išvados

- Darbe nustatyta, kad klavišų paspaudimo dinamika, kaip elgsenos biometrija, gali užtikrinti saugesnę ir universalesnę autentifikavimą, palyginti su tradicinėmis slaptažodžiais grindžiamomis sistemomis.
- Klavišų paspaudimų dinamikos transformavimas į vaizdus naudojant Gaboro filtro matricos transformaciją (GAFMAT) pagerino požymių išskyrimo ir gilaus mokymosi modelių tikslumą.
- Atliktų eksperimentų rezultatai parodė, kad siūloma kompleksinė autentifikavimo sistema sumažina vienodą klaidų lygį (angl. EER) ir padidina autentifikavimo tikslumą, taip patvirtindama sistemos efektyvumą nustatant anomalijas ir užkertant kelią neteisėtai prieigai.
- Daugiamatės įterpties vizualizavimo rezultatai pagerino sprendimų paramos galimybes vartotojo autentiškumo nustatymo naudojant klavišų paspaudimų dinamiką kontekste.

Apibendrinimas ir išvados

- GAFMAT metodo, skirto nevaizdiniams ar skaitiniams duomenims transformuoti į vaizdus, įdiegimas, kuris padidina duomenų panaudojimo universalumą ir išplečia jų taikymą gilaus mokymosi užduotims. Panaudojant šį metodą buvo parodytas ir pagrįstas jo efektyvumas tiek CMU, tiek GREYC-NISLAB duomenų rinkiniuose. CMU duomenų rinkinyje buvo gautas EER 0,04545, vidutinis tikslumas - 98,9 %. Panaudojus GREYC-NISLAB duomenų rinkinį buvo gautas ~98 % tikslumas ir EER svyravo nuo 0,044 iki 0,0755. Gauti rezultatai įrodo metodo efektyvumą atskiriant teisėtus naudotojus nuo potencialių įsilaužimų.
- Praplečiant kompleksinės autentifikavimo sistemos pritaikomumą, buvo apjungti CMU, KeyRecs, GREYC-NISLAB duomenų rinkiniai, pasitelkiant skirtingus interpoliavimo metodus. Gauti rezultatai rodo, kad tiesinė interpoliacija yra subalansuotas metodas, leidžiantis pasiekti mažiausią vidutinį EER ir pasižymintis stabiliu efektyvumu apjungiant skirtingus duomenų rinkinius.
- Naudojant konkrečias interpoliacija pagrįstas duomenų sujungimo strategijas, taip pat Siamo neuroninį tinklą su trigubų nuostolių funkcija, pasiektas geriausias EER nematytiems skirtingiems standartizuotiems duomenims buvo 0,13281. Tai rodo, jog šis metodas gali išplėsti naudotojų autentifikavimo sistemų galimybes, taip užtikrinant patikimesnes saugumo priemones kritinėms infrastruktūroms.

Apibendrinimas ir išvados

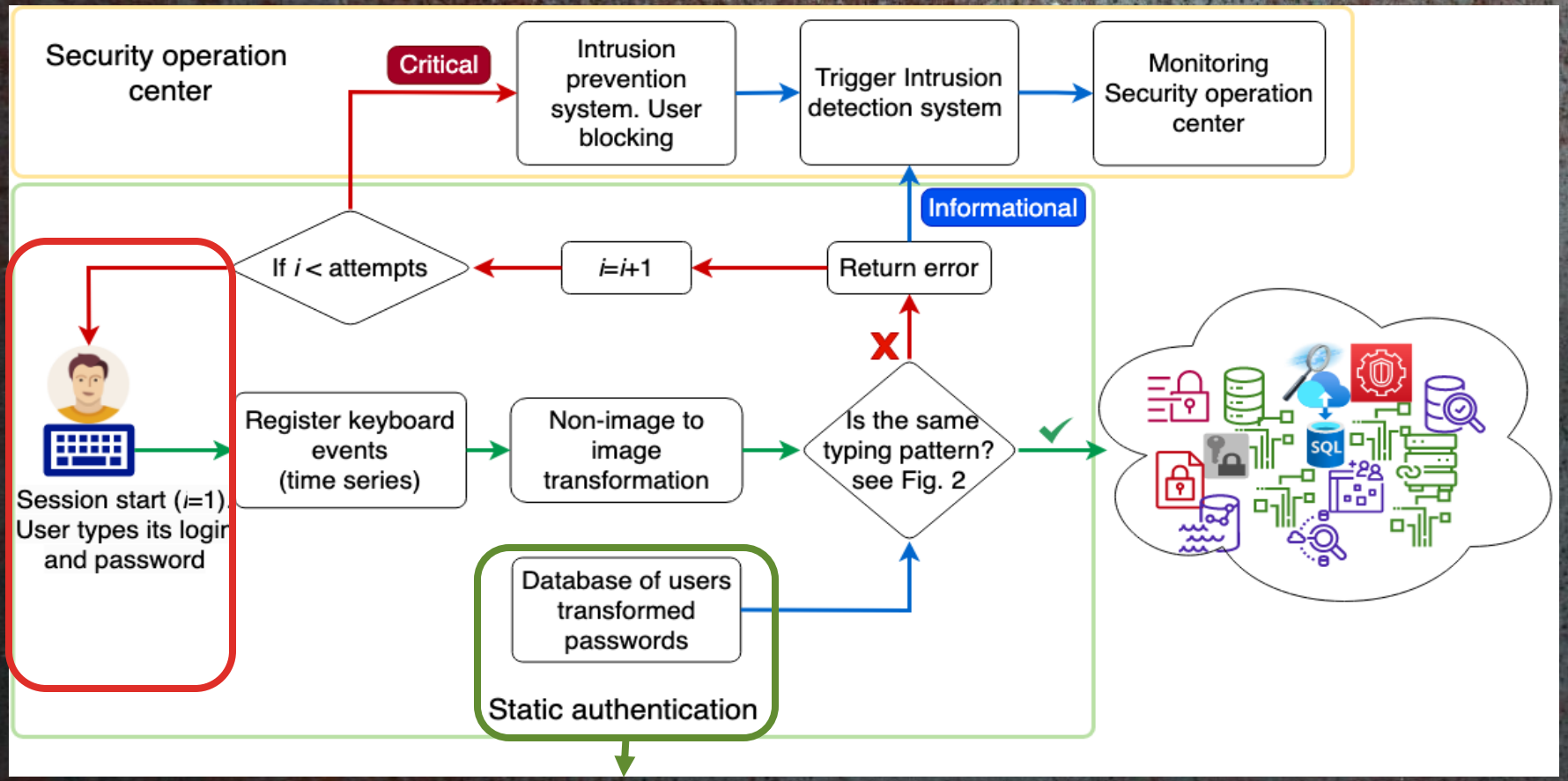
- Parengtas disertacijos rankraštis: Įvadas parengtas daugiau kaip 90 %, literatūros apžvalga - daugiau kaip 80 %. Empirinių rezultatų išbaigtumo lygis yra daugiau nei 70 %, išvados taip pat yra daugiau nei 70 % išbaigtos. Literatūros sąrašas taip pat atitinka reikalavimus, jo išbaigtumas yra daugiau kaip 80 %.

KITO PUSMEČIO DARBO PLANAS

- Baigti disertacijos rašymo procesą, pagerinti disertacijos kokybę pagal vadovo pastabas, gautas recenzentų pastabas.

- Daktaro disertacijos gynimas.

Vartotojų autentifikavimas



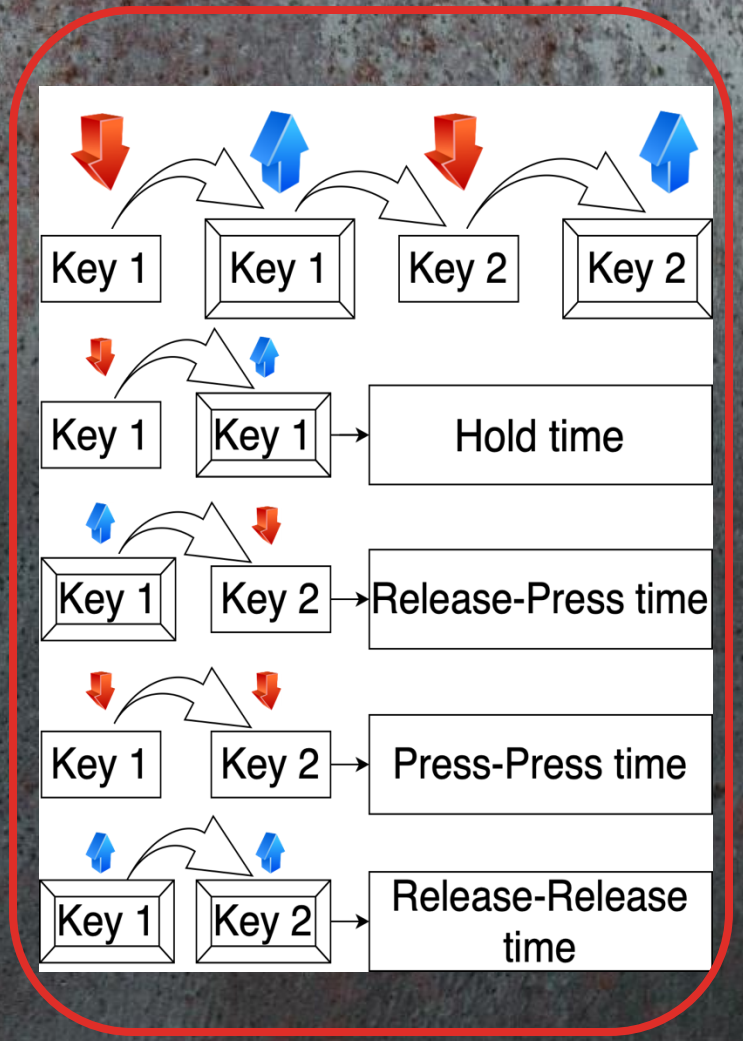
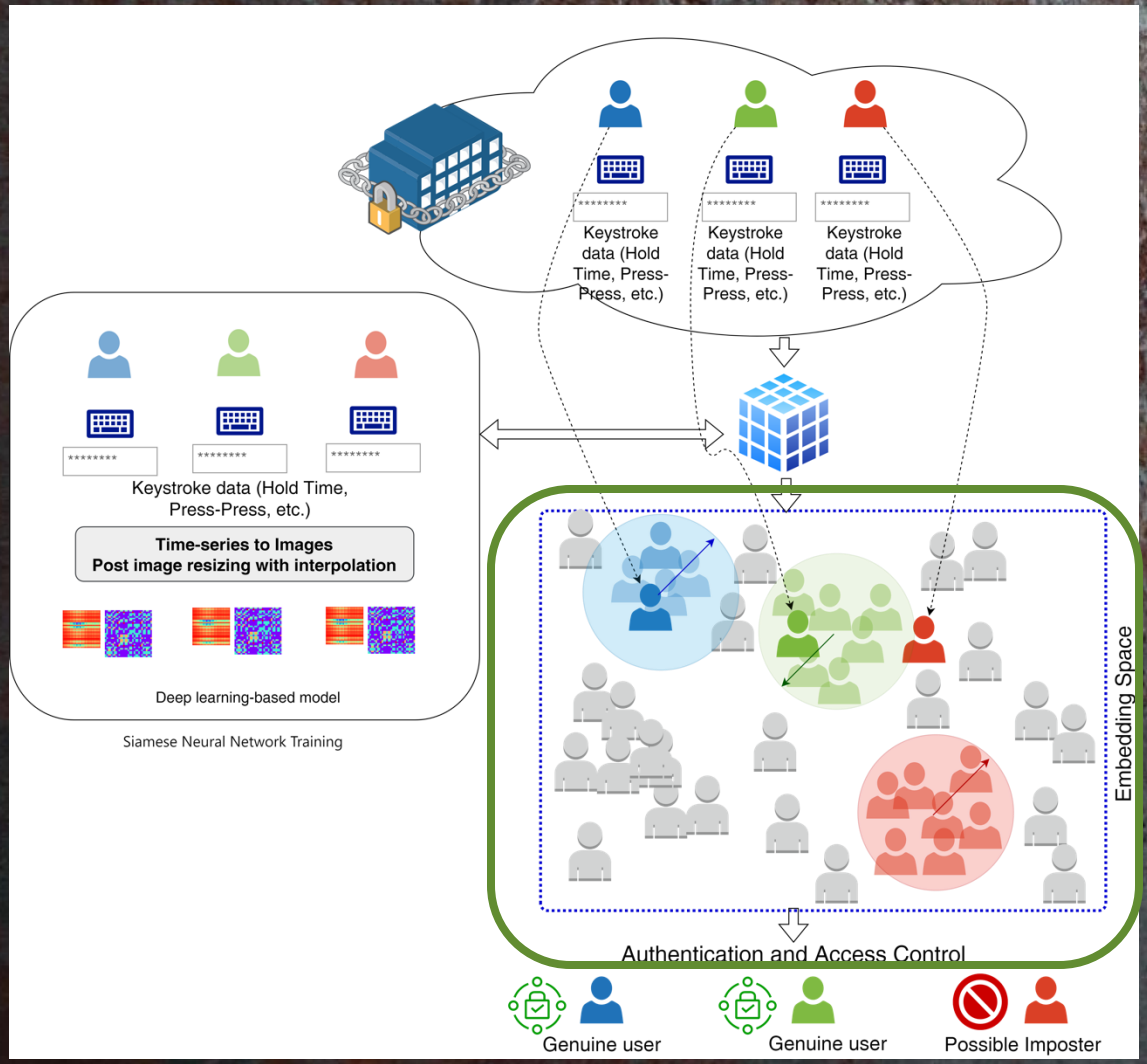


Fig. 1 Schematic representation of the user authentication process using an intrusion detection system and an intrusion prevention system based on user typing behavior.

VARTOTOJŲ AUTENTIFIKAVIMAS



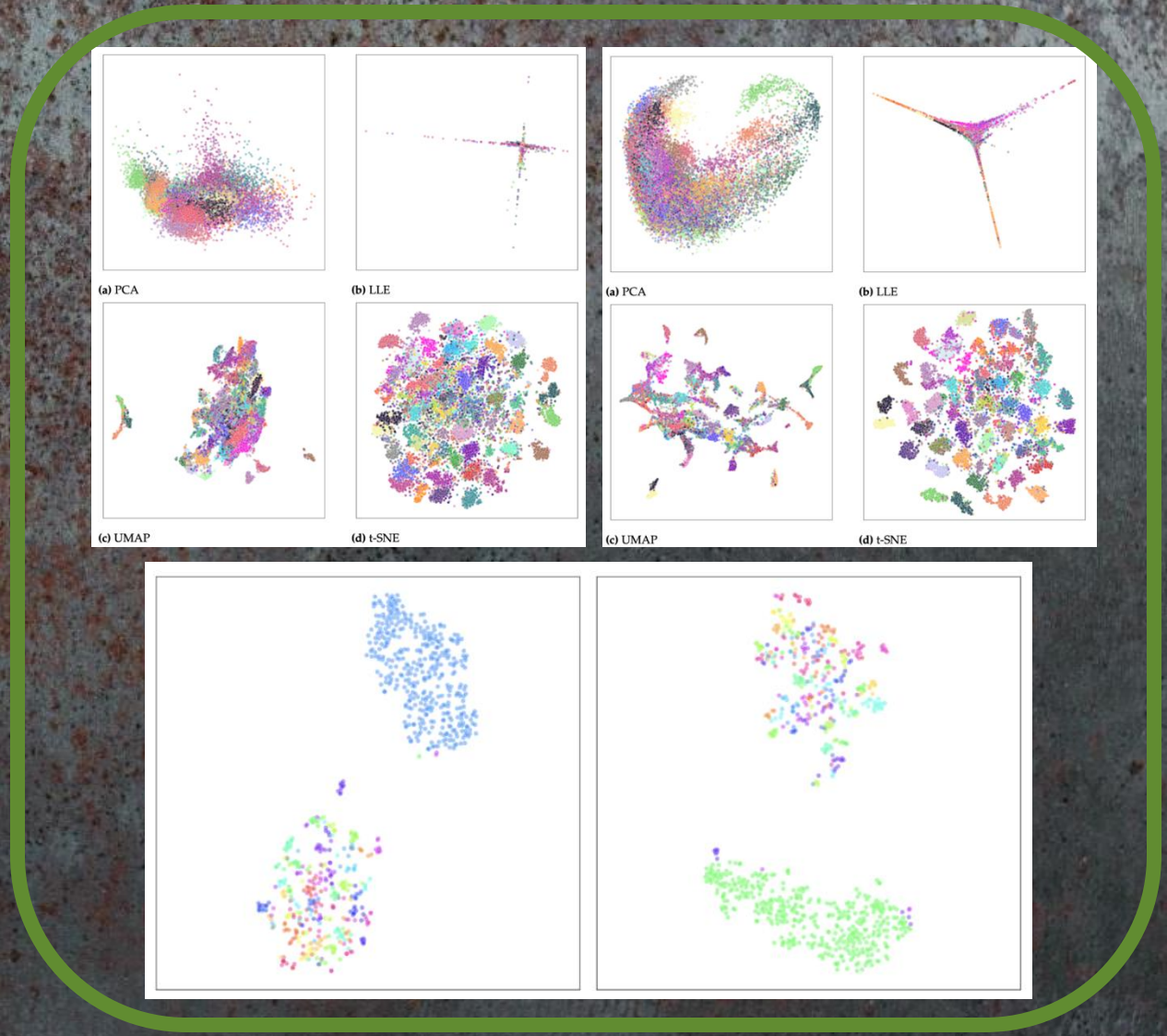


Fig. 2 Schematic representation of the user authentication process using an intrusion detection system and an intrusion prevention system based on user typing behavior.

Vartotojų autentifikavimas

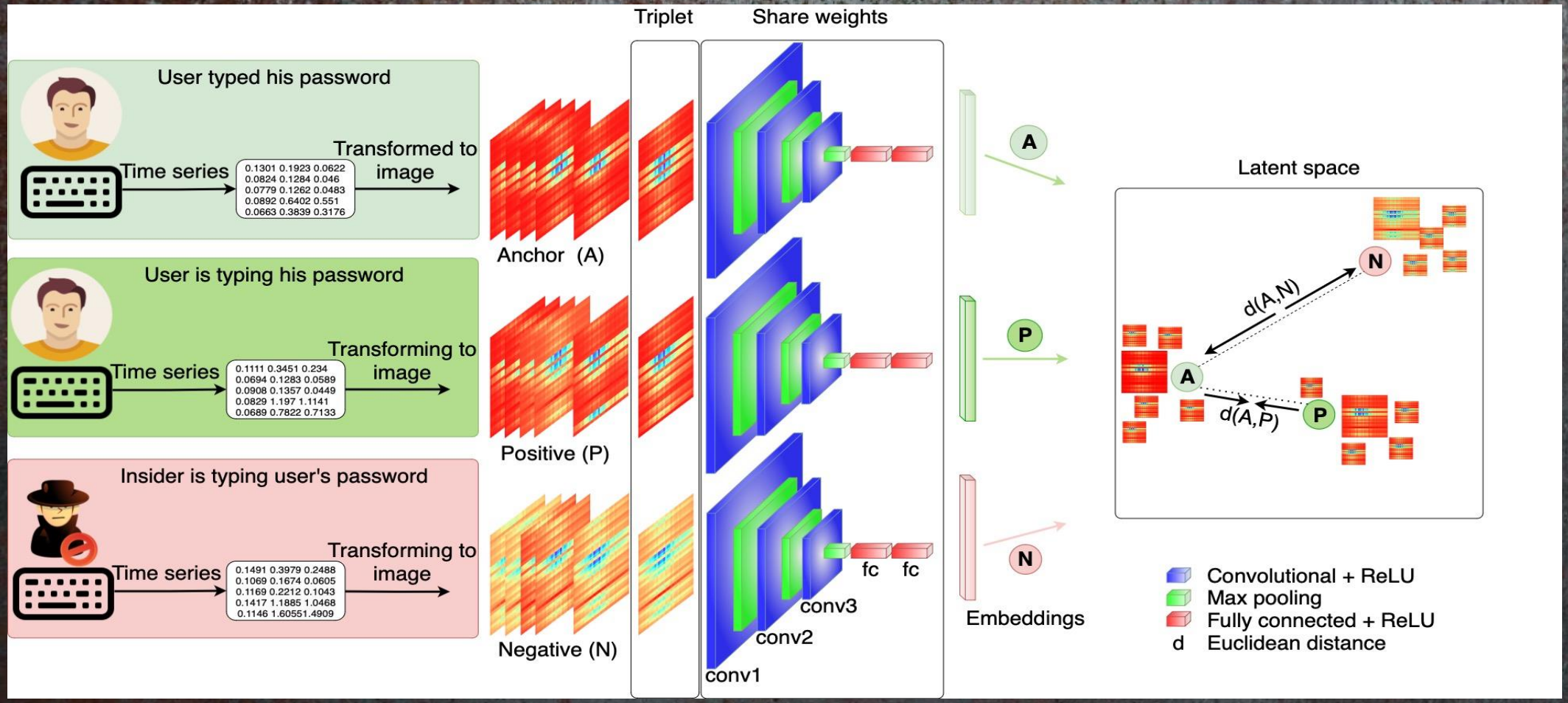


Fig. 3 Schematic representation of the proposed framework for time series transformation from keystroke biometric data features into images and training process of Siamese neural network with CNN branches

Gabor Filter Matrix Transformation

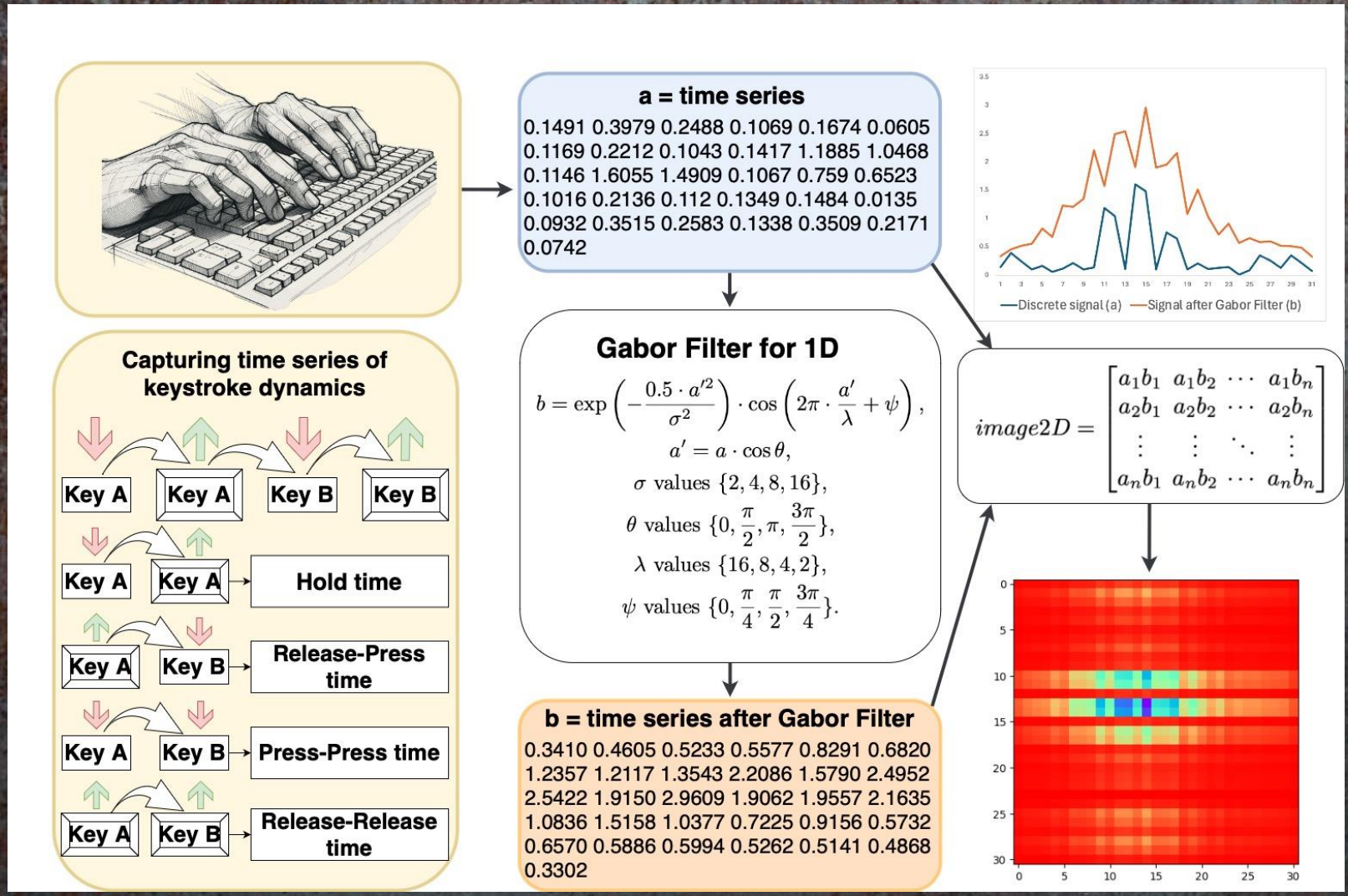


Fig. 4 Transforming the time series features of keystroke dynamics into an image using the GAFMAT approach: the process illustrates the application of Gabor filters to emphasize significant features in the data, followed by transforming the filtered data into a two-dimensional image that represents typing behavior

Duomenų suliejimas

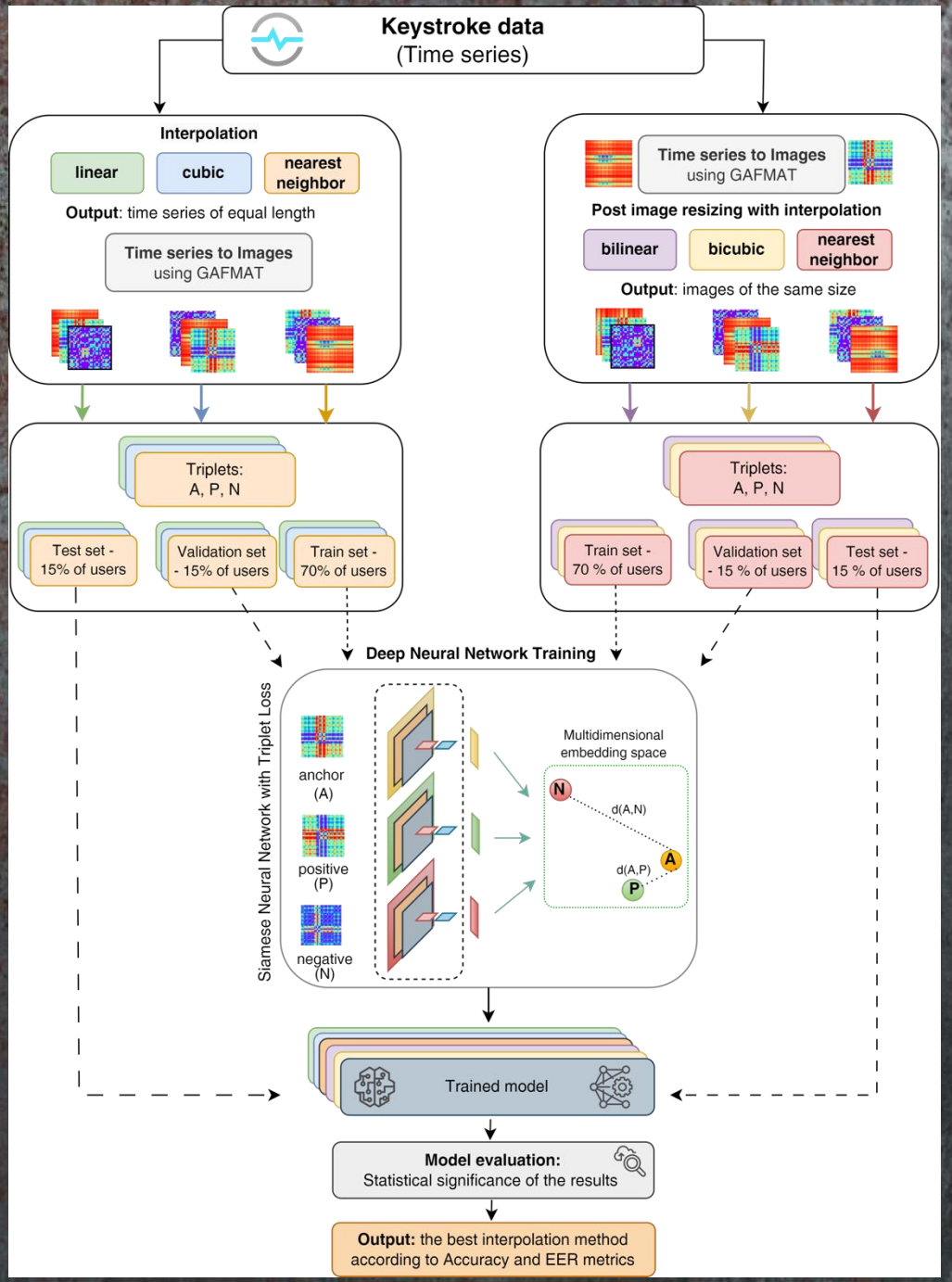


Fig. 4 Transforming the time series features of keystroke dynamics into an image using the GAFMAT approach: the process illustrates the application of Gabor filters to emphasize significant features in the data, followed by transforming the filtered data into a two-dimensional image that represents typing behavior

Duomenų suliejimas

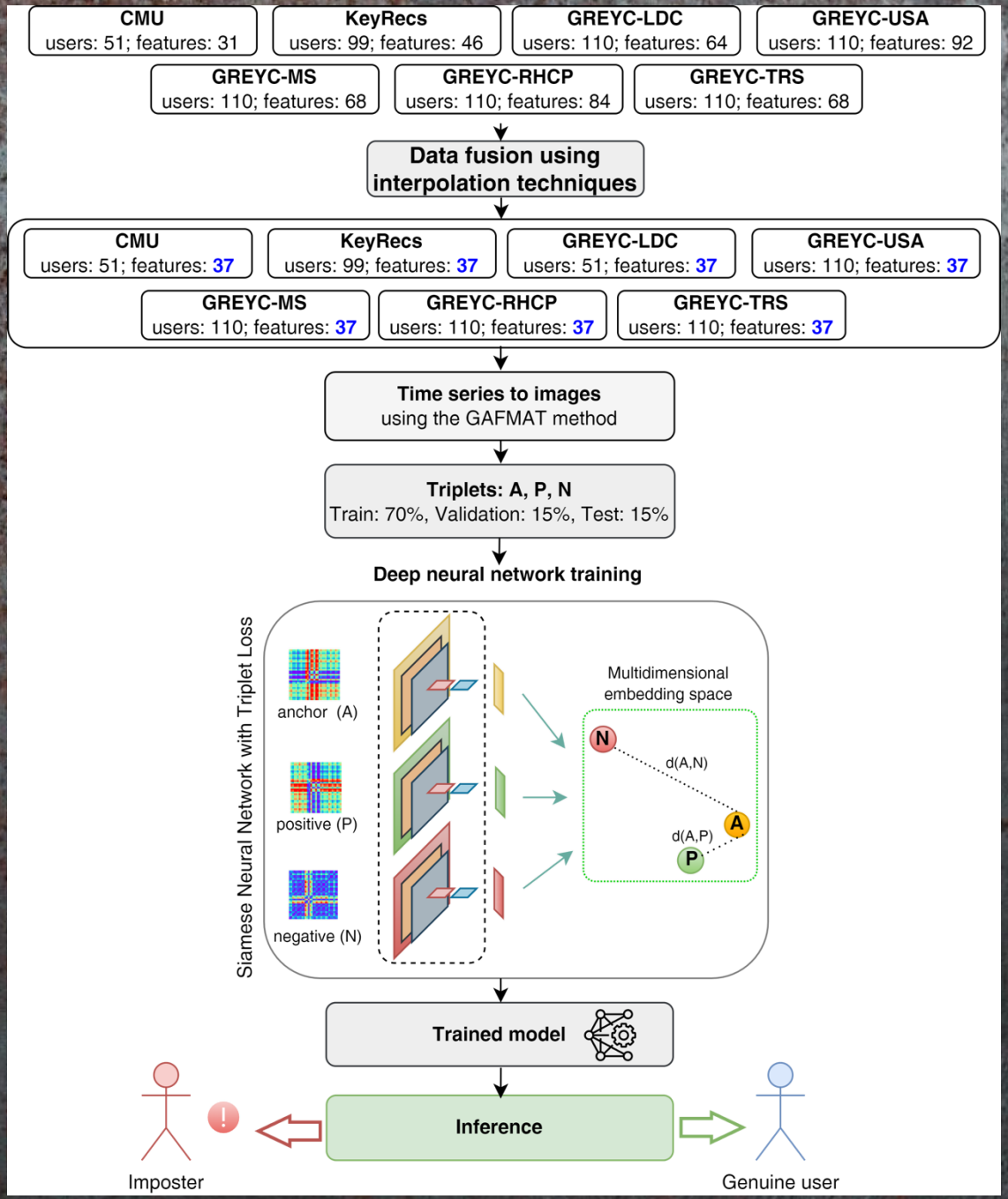
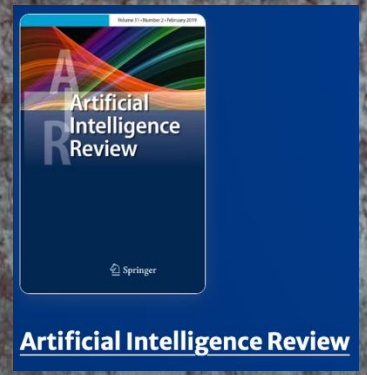


Fig. 5 Methodology for data fusion-based authentication using complex keystroke dynamics analysis: It includes steps for standardizing datasets through interpolation, transforming password samples into images, and using a trained Siamese neural network to compare embeddings of new inputs with stored records for verification

Rezultatai
GAFMAT palyginimas su kitais metodais



Metrics	Non-Image to Image Transformation Methods				
	GADF	GASF	RP	MTF	GAFMAT
Accuracy↑	0.99077	0.98473	0.98331	0.94744	0.98935
EER↓	0.04794	0.05540	0.05327	0.12074	0.04545
AUC↑	0.98612	0.98290	0.98394	0.94862	0.98668
AP_ED↓	0.44127	0.47255	0.43633	0.56487	0.48600
AN_ED↑	1.72784	1.71689	1.68884	1.59469	1.76378
AP_STD↓	0.27487	0.29295	0.28245	0.36906	0.31383
AN_STD↓	0.32888	0.34455	0.34881	0.40005	0.31295
AN_CS↓	0.45772	0.45264	0.46871	0.46011	0.43755
AP_CS↑	0.77936	0.76373	0.78183	0.71756	0.75700

Table 1 Results of image transformation methods on keystroke dynamics data from the CMU dataset using GADF, GASF, RP, MTF, and GAFMAT algorithms: Metrics-based evaluation on validation data.

Metrics	Passwords (GREYC-NISLAB)				
	leonardo dicaprio	the rolling stones	michaell schu-macher	red hot chilli peppers	united states of america
Accuracy↑	0.97656	0.98698	0.99219	0.97778	0.99220
EER↓	0.07552	0.04688	0.0651	0.04444	0.04688
AUC↑	0.97824	0.98667	0.98771	0.98272	0.98847
AP_ED↓	0.44736	0.43986	0.39958	0.45165	0.39566
AN_ED↑	1.55644	1.61202	1.48864	1.63478	1.61275
AP_STD↓	0.24318	0.21992	0.20467	0.21505	0.19676
AN_STD↓	0.40601	0.37381	0.38351	0.38917	0.38013
AN_CS↓	0.49905	0.48703	0.52795	0.47839	0.49790
AP_CS↑	0.77632	0.78007	0.80021	0.77417	0.80217

Table 2 Results using different accuracy metrics for passwords from GREYC-NISLAB on validation dataset when transforming time series features of keystroke dynamics into an image using the GAFMAT algorithm.

Rezultatai
Daugiamačių duomenų vizualizavimas

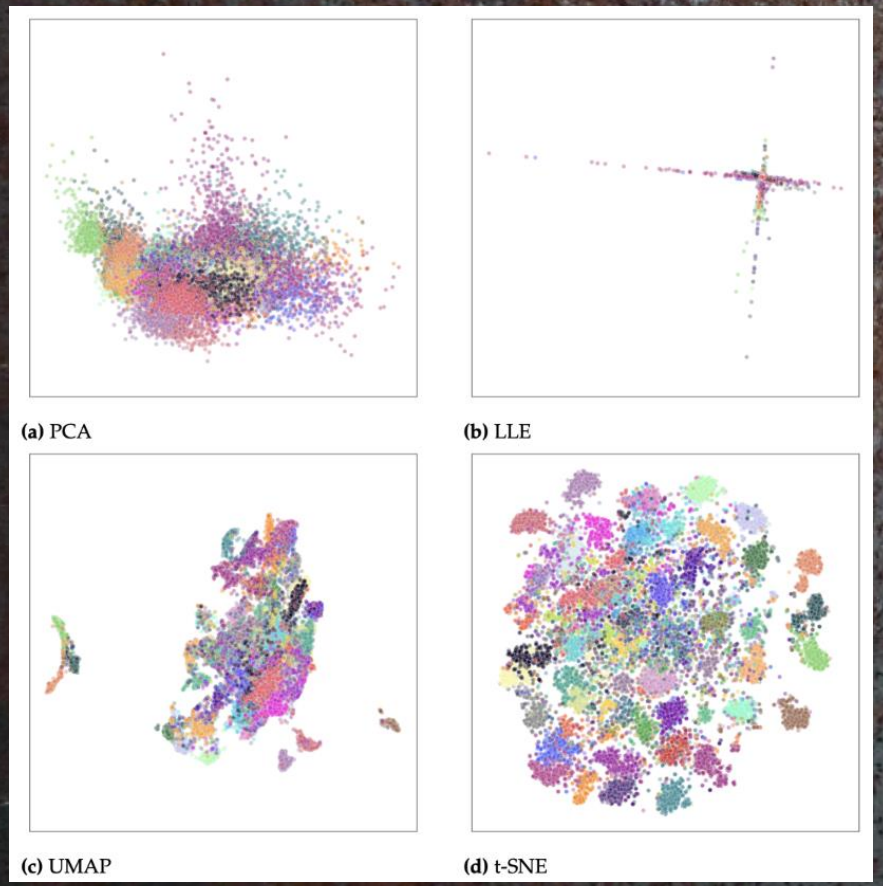


Fig 6. Multidimensional data visualizations by using different dimensionality reduction techniques: (a) PCA, (b) LLE, (c) UMAP, (d) t-SNE. Each color corresponds to a different user in the CMU dataset

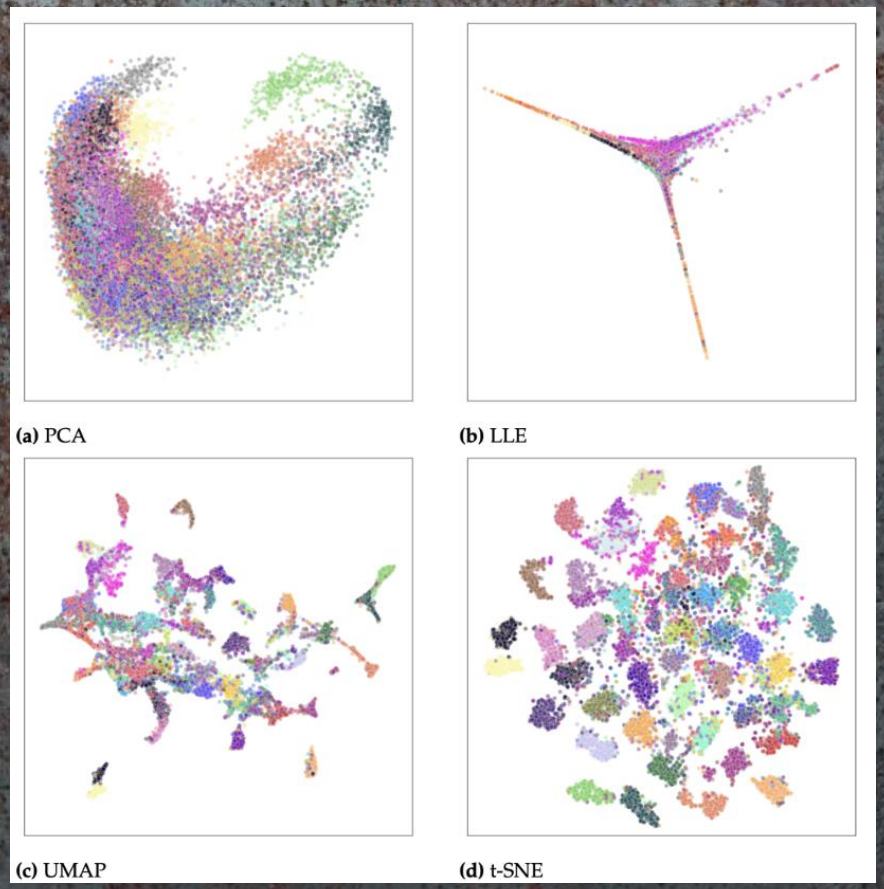


Fig 7. Visualization of multidimensional embeddings obtained by Siamese neural network using different dimensionality reduction techniques (p = 256): (a) PCA, (b) LLE, (c) UMAP, (d) t-SNE. Each color corresponds to a different user in the CMU dataset

Rezultatai
Daugiamačių duomenų vizualizavimas

Vizualizavimo rezultatų vertinimui pagal suformuotus klasterius buvo pasirinktas Silueto koeficientas.

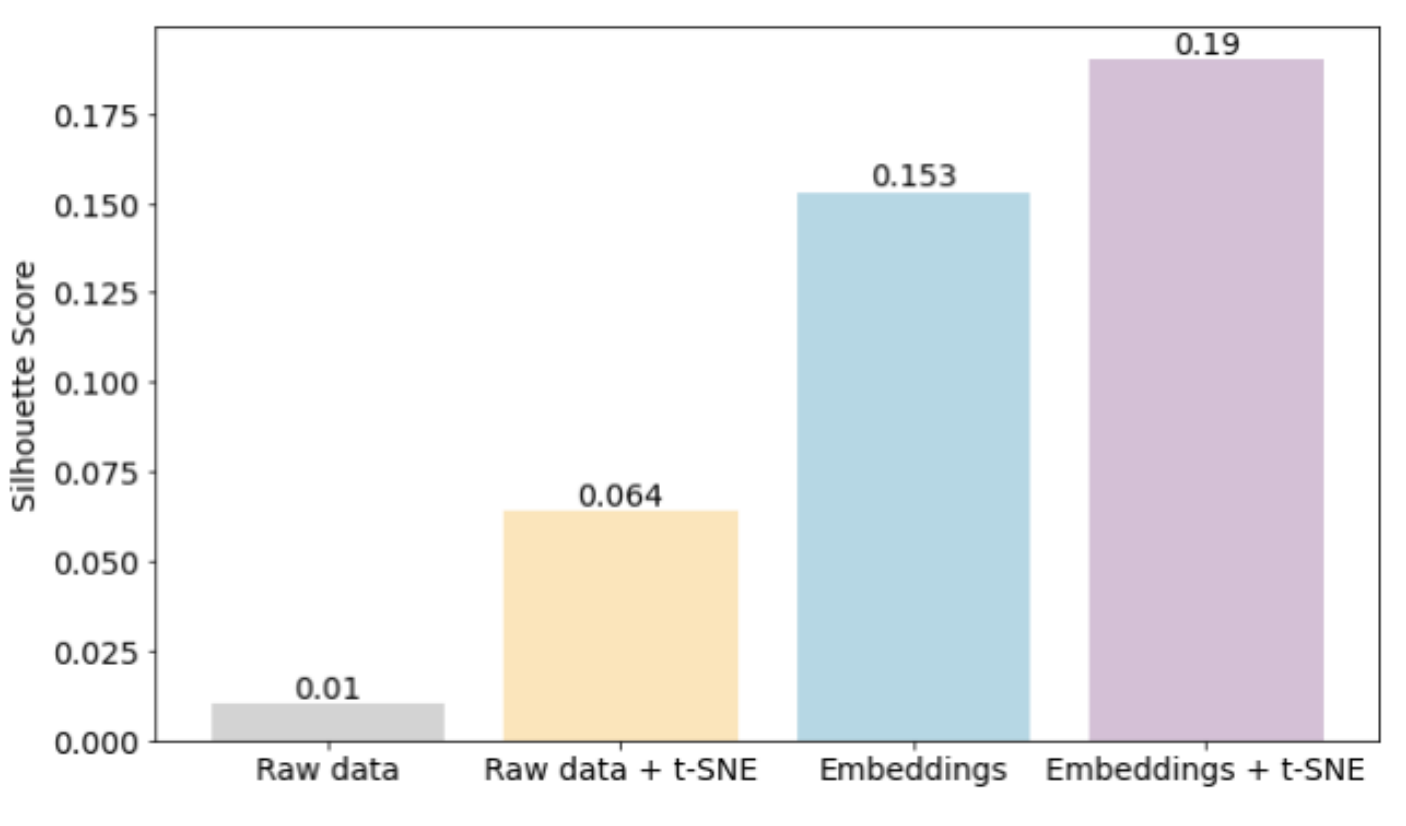


Fig 8. Silhouette scores before and after applying t-SNE on raw multidimensional data and their embeddings.

Rezultatai
Daugiamačių duomenų vizualizavimas

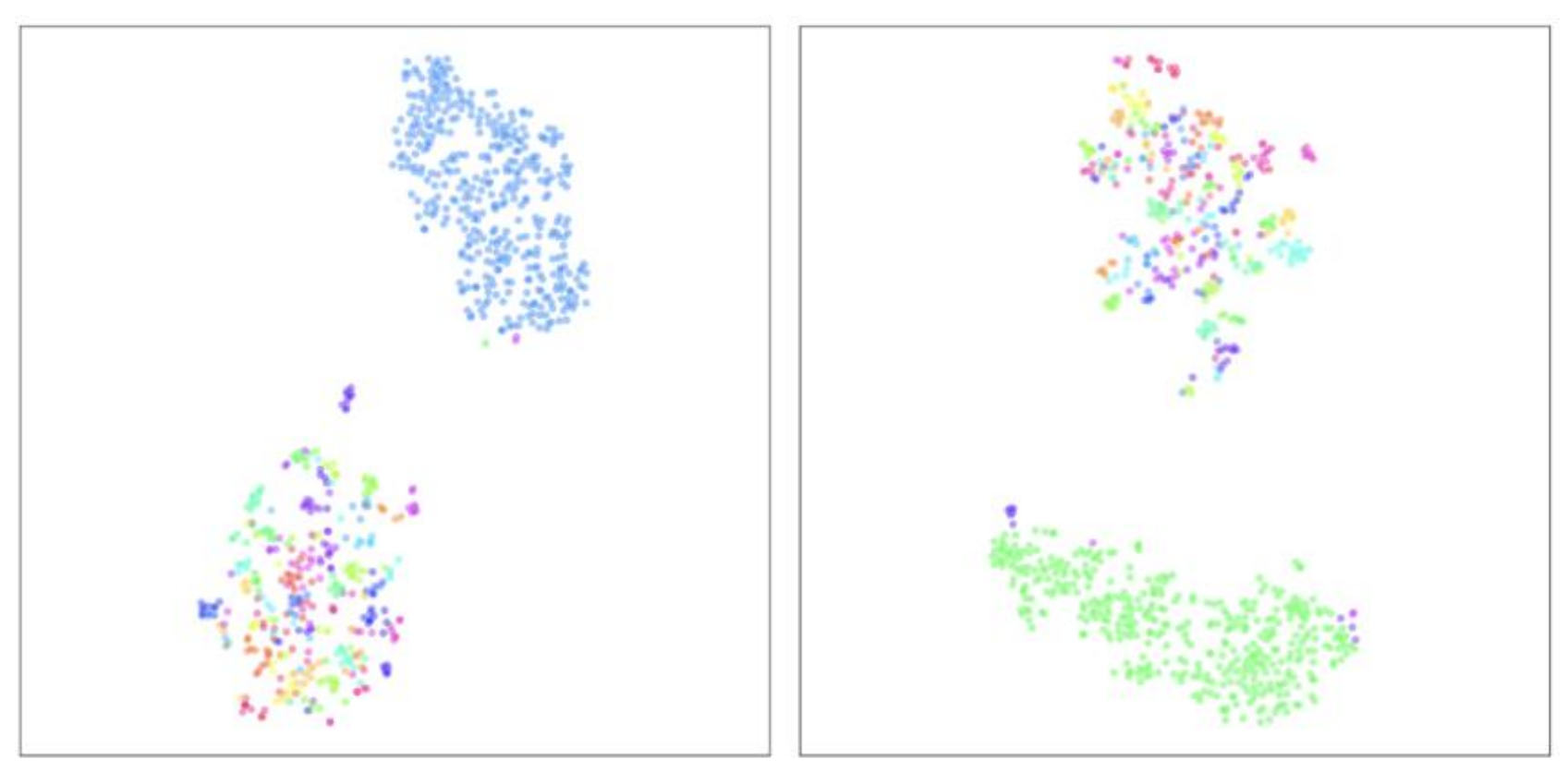
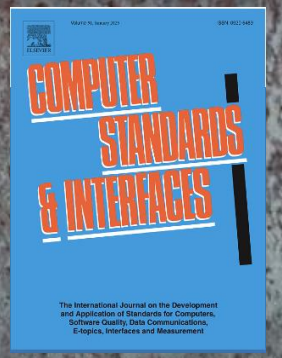


Fig. 9 Examples of visualizations that show password typing patterns of the same user and the other randomly selected users

Rezultatai
Duomenų suliejimas



	Time series interpolation			Post-resizing with interpolation		
Metrics	Cubic	Nearest Neighbor	Linear	Bicubic	Nearest Neighbor	Bilinear
Accuracy (best)	0.94141	0.92969	0.91797	0.91406	0.94141	0.92578
Accuracy (mean)	0.91685	0.91574	0.90737	0.90234	0.91016	0.91016
Accuracy (std)	0.01282	0.01062	0.00828	0.00835	0.01235	0.01772
EER (best)	0.15625	0.15625	0.13672	0.15234	0.16016	0.16016
EER (mean)	0.17634	0.17913	0.16462	0.18025	0.17187	0.17188
EER (std)	0.01423	0.02007	0.01423	0.01376	0.01023	0.02738

Table 3 Performance measures using the CMU dataset using different interpolation methods for time series data and image post-resizing

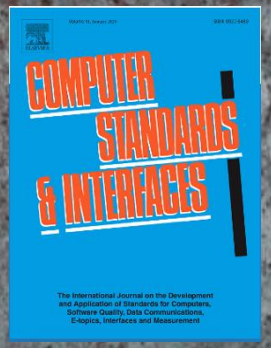
	Time series interpolation			Post-resizing with interpolation		
Metrics	Cubic	Nearest Neighbor	Linear	Bicubic	Nearest Neighbor	Bilinear
Accuracy (best)	0.91406	0.90625	0.91406	0.89844	0.91016	0.90625
Accuracy (mean)	0.87956	0.89388	0.89779	0.88346	0.89388	0.89388
Accuracy (std)	0.01648	0.01264	0.01569	0.01159	0.01137	0.00941
EER (best)	0.21484	0.18359	0.15625	0.18750	0.15625	0.17188
EER (mean)	0.21940	0.20833	0.19401	0.20443	0.18424	0.19141
EER (std)	0.00417	0.01909	0.02296	0.01120	0.01633	0.01256

Table 4 Performance measures using the KeyRecs dataset using different interpolation techniques

	Datasets				
Metrics	LDC	MS	RHCP	TRS	USA
Accuracy (best)	0.93750	0.96875	1.00000	0.96875	0.90625
Accuracy (mean)	0.87500	0.93229	0.87500	0.89583	0.83854
Accuracy (std)	0.03608	0.04199	0.06250	0.03898	0.04199
EER (best)	0.15625	0.15625	0.15625	0.1250	0.15625
EER (mean)	0.19792	0.18750	0.21354	0.17708	0.22396
EER (std)	0.02329	0.02552	0.04199	0.02946	0.03335

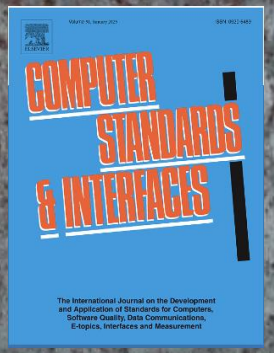
Table 5 Performance measures using the GREYC-NISLAB datasets using linear interpolation technique

Rezultatai
Duomenų suliejimas



- Pirmajame eksperimente slaptažodžių frazės iš KeyRecs, CMU ir GREYC-NISLAB duomenų rinkinių buvo sujungtos į vieną duomenų rinkinį, kad būtų galima apmokyti SNN modelį. Apmokytas modelis buvo įvertintas naudojant nematytą pavyzdžių poaibį, kurį sudarė 15 % naudotojų iš kiekvieno iš šių duomenų rinkinių.
- Antrasis eksperimentas buvo skirtas pastebėtoms efektyvumo problemoms, susijusioms su ilgesnėmis slaptažodžių frazėmis, spręsti. Šiuo atveju KeyRecs ir GREYC-NISLAB duomenų rinkiniai buvo sujungti, neįtraukiant ilgesnių RHCP ir USA slaptažodžių frazių, nes buvo pripažinta, kad ankstesnėse analizėse jų rezultatai buvo prastesni. Apmokytas modelis buvo įvertintas naudojant visus CMU duomenis.
- Trečiajame eksperimente į KeyRecs, CMU ir GREYC-NISLAB duomenų rinkinių susiliejimą nebuvo įtrauktos RHCP ir JAV slaptažodžių frazės, darant prielaidą, kad pašalinus ilgesnes slaptažodžių frazes gali pagerėti bendras efektyvumas. Po to apmokytas modelis buvo išbandytas su nematytu 15 % naudotojų iš CMU duomenų rinkinio.

Rezultatai
Duomenų suliejimas



- Metodikos efektyvumui įvertinti buvo taikoma ANOVA metodas, skirtas statistiškai įvertinti skirtingų interpoliavimo metodų veiksmingumą. Atlikus analizę gautos p-reikšmės viršijo 0,05 ribą, o tai rodo, kad interpoliavimo metodų veiksmingumas įprastu reikšmingumo lygmeniu statistiškai reikšmingai nesiskiria.
- Naudojant konkrečias interpoliacija pagrįstas duomenų sujungimo strategijas, taip pat Siamo neuroninį tinklą su trigubų nuostolių funkcija, pasiektas geriausias EER nematytiems skirtingiems standartizuotiems duomenims buvo 0,13281.

Klausimai?