

Feasibility of Conditional Variational Autoencoders for Phase-Averaged Synthetic Time Series

Matthias Rüb¹, Jens Grüber¹, and Hans Schotten^{1,2}

¹Intelligent Networks Research Group, German Research Center for Artificial Intelligence, Kaiserslautern, Germany

²Institute for Wireless Communication and Navigation, RPTU Kaiserslautern-Landau, Kaiserslautern, Germany

Matthias.Rueb@dfki.de

Jens.Grueber@dfki.de

Schotten@rptu.de

Abstract: In cybersecurity, synthetic data is beneficial for testing, training, and enhancing AI-driven defense systems without compromising sensitive information. Critical sectors like telecommunications, finance, energy, and healthcare generate vast amounts of time-series data, often requiring reduction methods such as phase-averaging to manage scale. However, this can obscure essential features, impacting anomaly detection and threat modeling. This study explores whether conditional Variational Autoencoders (cVAEs) can generate high-quality synthetic data when given only phase-averaged time series for training. Results on a biometric use-case show that cVAEs preserve intrinsic properties of reduced data, making it usable for classification and to a more restricted degree as training data in downstream cybersecurity applications.

Keywords: Synthetic data, Biometric security, Soft biometrics, Data reduction

1. Introduction

Synthetic time series data is becoming increasingly important for cybersecurity applications as AI models grow more capable of processing large amounts of multimodal data (Agrawal, Kaur and Myneni, 2024). This is particularly relevant with recent AI models, which can enhance automation in threat detection, threat modeling, and initiating mitigation strategies (Sowmya and Mary Anita, 2023). For those use-cases synthetic data can aid in modeling rare special cases—a common weakness of AI models—and improve model generalization. It can also help address deficiencies in available datasets, such as imbalances and the scarcity of certain classes. Biometric data, in particular, is often scarcely available and subject to strict data protection regulations, which hampers the development of AI-based cybersecurity models (Rüb *et al.*, 2022). As the computational complexity of AI models continues to increase, it becomes necessary to apply data reduction methods (Reddy *et al.*, 2020). This raises the question of whether during synthetic data generation, sufficient information from time series data can be retained without the use of even more complex models.

This work investigates whether a simple synthetic generator with low computational complexity can capture enough intrinsic information about individuals from phase-averaged soft-biometrics to allow a classifier to identify the person correctly. The exemplary biometric case study involves gait analysis data collected with a smart insole featuring pressure-sensitive sensors. The dataset includes 30 participants, walking at a steady pace. Evaluations using various classification schemes show that a low-complexity architecture of a conditional variational autoencoder (cVAE) is indeed capable of generating sufficient information from phase-averaged time series data to enable correct identification, even with heavily reduced data of 8 points per gait cycle within the limited dataset of 30 individuals. Further difficulties that the cVAE model has to overcome in this exemplary case study are a small data set, a small latent space with 2 dimensions, only a few trainable parameters for the cVAE network, and a non-linear time axis due to the phase-averaging process. The rest of this work is structured as follows: Section 2 puts this work into context of related work. Section 3 describes the method used and gives information about the dataset. In section 4 the experimental results are presented and analyzed with an explanation of the metrics on a tutorial level. It is shown that even with severe data reduction, the low complexity cVAE model is able to generate synthetic biometric data. Section 5 concludes this work and highlights possible future research directions.

2. Related Work

Due to advancements in artificial intelligence synthetic data has grown more relevant in cyber security and biometric applications: (Osadchy *et al.*, 2017) have described the role of synthetic face generation for cyber security solving the problem of small data sets in the training of ai models. Similar to this work, (Papavasileiou *et al.*, 2021) use gait-based authentication. The authors fuse accelerometer data with ground contact force data.

In contrast, this work only utilizes the force data similar to (Herbst *et al.*, 2024) and even reduces it with phase averaging, limiting the available information about the different classes.

For synthetic data generation, various approaches are discussed in the literature, which generally entail high computational complexity. Generative Adversarial networks (GANs) proposed by (Goodfellow *et al.*, 2014) and diffusion models for synthetic time series (Kong *et al.*, 2021) are capable of generating high quality synthetic time series. In the case of GANs they do not only require large computational efforts but require sophisticated fine tuning. This work utilizes cVAEs proposed by (Sohn, Lee and Yan, 2015). This expansion of the initial VAE proposed by (Kingma and Welling, 2014) allows the synthetic generator to create data of specific labels. In contrast to the previously mentioned methods, the VAE structure requires lower computational complexity. This method is widely used in the context of biometric data and time series imputation (Fortuin *et al.*, 2020).

For the evaluation of synthetic time series, the scientific community still has not agreed on unified criteria. The following evaluation schemes are widely popular: In the domain of image generation (Heusel *et al.*, 2017) proposed Fréchet Inception Distance (FID) as a performance metric for synthetic generators. It compares the features of an artificial neural network of synthetic and real data but requires a pre-trained classification network. Further, multiple adaptations of this metric for time-series have been proposed (Lee, Malacarne and Aune, 2023). (Esteban, Hyland and Rättsch, 2017) proposed the evaluation of synthetic data based on the performance of downstream classifying tasks performed by an arbitrary machine learning model. As the classification with artificial intelligence is the crucial element for cyber security applications the metrics proposed by (Esteban, Hyland and Rättsch, 2017) are utilized in this work.

3. Methods, Synthetic Data Generation and Metrics

This work utilizes a dataset containing five minutes of consistent walking data collected from 30 participants. All participants were aged between 17 and 29 years old. Both male and female individuals are included in the study. The raw data is obtained with a pressure-sensitive insole collecting data every ~20 ms. The pressure is measured and summarized over 12 different areas on the insole according to ('Stappone', 2025). The raw data is normalized. Note that this partially eliminates the weight of the person as a biometric feature, however indirect features that relate to the weight are still present, e.g. the weight distribution over the insole during a gait cycle.

The raw data is further separated into 8 distinct phases. A detailed description of the phases and their physiological interpretation is given in (Feiste, 2004). In general, the first part of the gait cycle can be interpreted as a stance phase which contributes ~60 % to the full cycle. The remaining ~40 % are attributed to the swing phase. These main phases are further separated into 8 smaller phases described in Table 1.

For the identification process, the Rocket method proposed by (Dempster, Petitjean and Webb, 2020) is utilized and competitive in the state-of-the-art time series classification (Middlehurst *et al.*, 2021). This method applies random convolutional kernels along the time series data, which given a large enough kernel size, recognizes enough relevant features for persona identification requiring low computational complexity. For this work, the time series is extremely short with only 8 time points after phase averaging. The kernel size is chosen as 500. Irrelevant features do not impact the performance of the default Ridge Classifier further described in (Singh, Prakash and Chandrasekaran, 2016).

Table 1: Phases of a gait cycle

Phase	Part of the full cycle
1	0 – 2 %
2	2 – 10 %
3	10 – 30 %
4	30 – 50 %
5	50 – 60 %
6	60 – 73 %
7	73 – 87 %
8	87 – 100 %

For evaluation of the performance of the cVAE, the evaluation scheme of (Esteban, Hyland and Rättsch, 2017) is utilized. It was initially created for the evaluation of generative adversarial networks (GANs). The first method “Train on Synthetic Test on Real (TSTR)” uses a synthetically generated dataset to train a classification model. This model is tested on a held-out test dataset of real samples. Therefore, the intrinsic properties of the synthetic data must be distinctive enough for the classification model to transfer this learning for usage on real data. The requirement of labels for this method is fulfilled by the cVAE as this per definition creates labelled data.

The second method as described in (Esteban, Hyland and Rättsch, 2017) is the “Train on Real Test on Synthetic (TRTS)” scheme: A classification model is trained on real data and afterwards tested on synthetic data. A disadvantage of this method is that it does not punish the failure modes of the synthetic generator of just copying real data or a lack of diversity in the generated samples. However, the metric is a valuable indicator for other failure modes of synthetic generation, for instance no convergence during the training process and therefore very low-quality samples.

The model used is based on the concept of a conditional Variational Autoencoder (cVAE), a refined version of the variational autoencoder approach, which incorporates the label information into both encoding and decoding, as described by (Sohn, Lee and Yan, 2015). A schematic overview is given in Figure 1. This experimental study has 30 labels, each corresponding to a single participant in the original dataset. This information is introduced into the network via a concatenation layer which merges the input data with the corresponding label.

The Encoder Network consists of a fully connected dense layer, followed by two dense layers for mean and standard deviation calculation. Aided by this encoder network, a compressed representation of a given input is embedded into the 2-dimensional latent space. The decoder utilizes this representation, combined with a given condition, to generate new data. The latent space is modeled with a gaussian distribution, meaning points sampled closer to the mean yield generated data that more closely resembles the original input. Based on the encoder characteristics the decoder consists of a concatenation layer, and two consecutive dense layers, establishing a network capable of effectively decompressing the latent space embeddings and therefore constructing new data.

The dense layers in both the encoder and decoder networks each consist of 20 nodes. The entire cVAE network used in this work comprises a total of 2020 trainable parameters. Note that this network is extremely small (IoT-level). Due to the small dataset, a larger network would be more prone to overfitting.

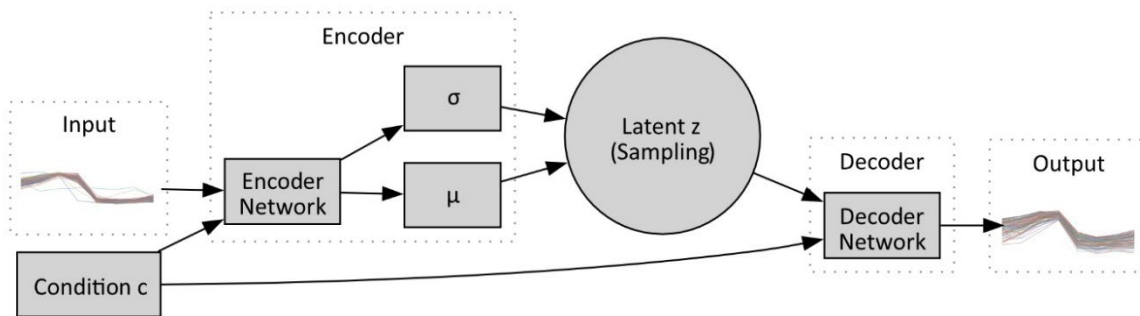


Figure 1: Block Diagram of a conditional Variational Autoencoder with layers for mean (μ) and standard deviation (σ)

As previously outlined the study presented in this work incorporates step data of 30 individuals walking for a fixed duration. The dataset serves as a baseline to generate new synthetic datasets. These steps are evenly divided between the cVAE and the classifier. While the dataset for the cVAE is split into training, validation, and test sets (80%, 16%, and 4%, respectively), the rocket classifier has a training and test dataset in a 0.5 split. To compensate for the non-deterministic behavior of a cVAE-network the training and data process was rerun a total of 60 times. To ensure comparability across experiments, each trial consists of 10 epochs with a constant batch size of 8. The synthetically generated datasets were classified using a separate instance of the classifier, which was retrained for each iteration. The original dataset was randomly shuffled and distributed evenly between the cVAE and rocket. Note, that due to the separation of the dataset for the classifier and synthetic generator, some intrinsic properties of the different classes can be captured by the cVAE even through copying real data. Therefore the TSTR metric can appear higher, than the true capability of the generator to extract distinct features.

As a first step, time series classification is used to confirm that the phase-averaged gait data contains sufficient intrinsic information about each individual for identification. Only if this fundamental requirement is met can the synthetic data reproduce these intrinsic features in a class-specific manner. For this purpose, the Rocket classifier by (Dempster, Petitjean and Webb, 2020) is used. The smart insole pressure data is a multivariate time series with 12 sensors distributed over the insole area. For further data reduction, only two of these 12 sensors are utilized for the machine learning-based classification: One sensor area at the outer toe area and one in the inner part of the middle of the insole. These two sensors are chosen based on the feature permutation importance, therefore these two have a high impact on classification accuracy. The median accuracy of the Rocket classifier, with a train-test split of 50 % of the real data, ranges from 0.45 to 0.96 for the different classes as shown in Figure 2.

4. Experimental Results

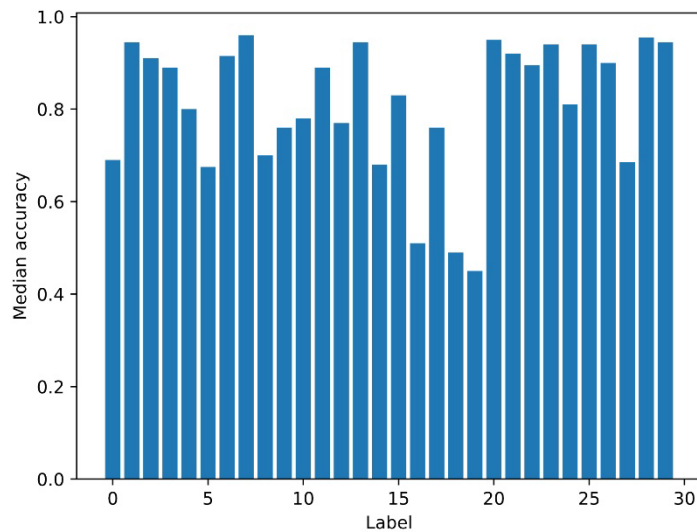


Figure 2: Median accuracy of Rocket time series classification on a single gait cycle utilizing only 2 out of 12 sensors. The median is calculated from 60 trials, each shuffling the train-test split of the data. The class-labels are indexed from 0 to 29

The differences may, for example, lie in the fact that some participants stand out due to their body size or have a particular foot shape, which can also influence the classification alongside walking behavior. This discrepancy between the subjects in the classification of real data must be considered when evaluating synthetic data.

Exemplary real pressure data of a gait cycle are shown in Figure 3a. The exemplary sensor for this representation is located in the heel area of the insole. At the beginning of the gait cycle, the measured pressure is therefore high and decreases as the foot rolls off. Additionally, three outlier graphs can be identified in the diagram, which do not follow the standard deviation of the other measurements. These may result from isolated errors in the electronics of the sensor system.

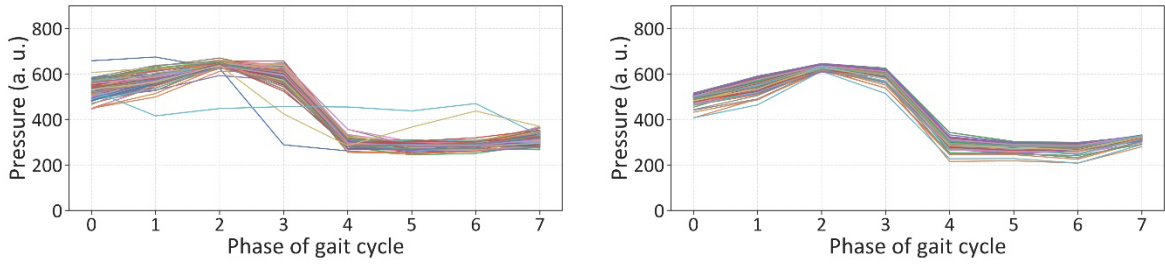


Figure 3a,b: Sample of one gait cycle from the original dataset (left) and synthetic step with standard deviation (σ_{dec}) of 3 of the gaussian latent sampling (right)

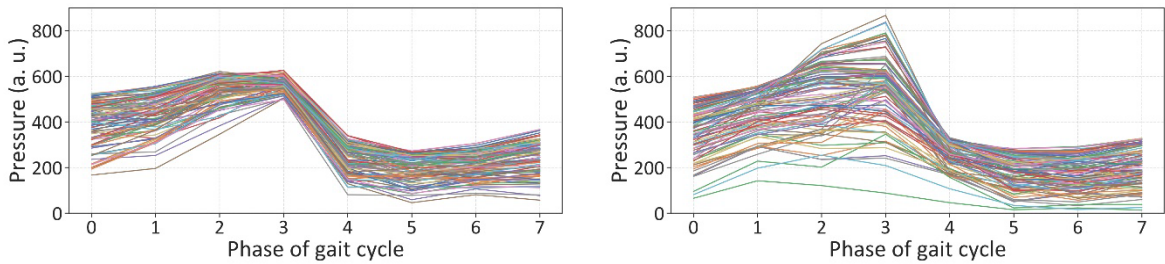


Figure 4a, b: Generated steps with standard deviation (σ_{dec}) of 8 and 13 of the latent sampling respectively

The synthetic generation process by the cVAE is performed according to Section 3 through latent sampling of a Gaussian distribution with a mean of (0,0) and a standard deviation of the decoder, σ_{dec} ranging from 1 to 15. Since a Gaussian distribution is also used during the training process, latent values close to the mean tend to be overtrained; consequently, the synthetic generator frequently simply copies real data in this range. When sampling from regions further out in the latent space, the resulting synthetic data show greater variation. This is illustrated in Figure 4a,b with standard deviations σ_{dec} of 8 and 13, respectively. These exhibit significantly greater variation and, thus, a larger deviation from the real data shown in Figure 4a. This qualitative, visual assessment aligns well with the TRTS metric presented in Figure 5.

The diagram shows that as the standard deviation of latent sampling increases, the TRTS evaluation metric decreases. According to the definition of the TRTS metric in Section 3, an ML classifier is first trained on real data and then tested on synthetic data. However, this metric can yield good results even when the synthetic generator we want to evaluate simply copies the real data. Consequently, the cVAE already provides qualitatively good results, but it is unclear whether it merely copies real data or captures intrinsic properties of the data and synthesizes them as new time series.

To address this, a second metric is used to evaluate the synthetic data: the TSTR metric, also shown in Figure 5. For this metric-a Rocket classifier is first trained on synthetic data and then tested on real data. The classification of real data is only possible if the synthetic time series contain sufficient features to distinguish the time series of one class from others. However, it is only relevant that the classification algorithm detects abstract patterns, which do not necessarily align with human evaluation of the synthetic data quality.

The diagram in Figure 5 shows that TSTR initially increases with increasing sampling standard deviation, reaching approximately 0.85, and then plateaus. The classification accuracy thus remains high, even though the samples no longer visually match real data, as shown in Figure 4b. For the evaluation of synthetic data, both metrics are therefore relevant, as the data should both correspond with real data by visual inspection (mostly aligned with TRTS) and contain intrinsic properties of the individuals being classified (TSTR). At the intersection of the two curves, where TRTS decreases and TSTR increases, both metrics reach a value of 0.80. This puts the evaluation metrics of the synthetic data only slightly below the comparison with real data, where 60 trials using the Rocket classifier achieve a median classification accuracy of 0.88.

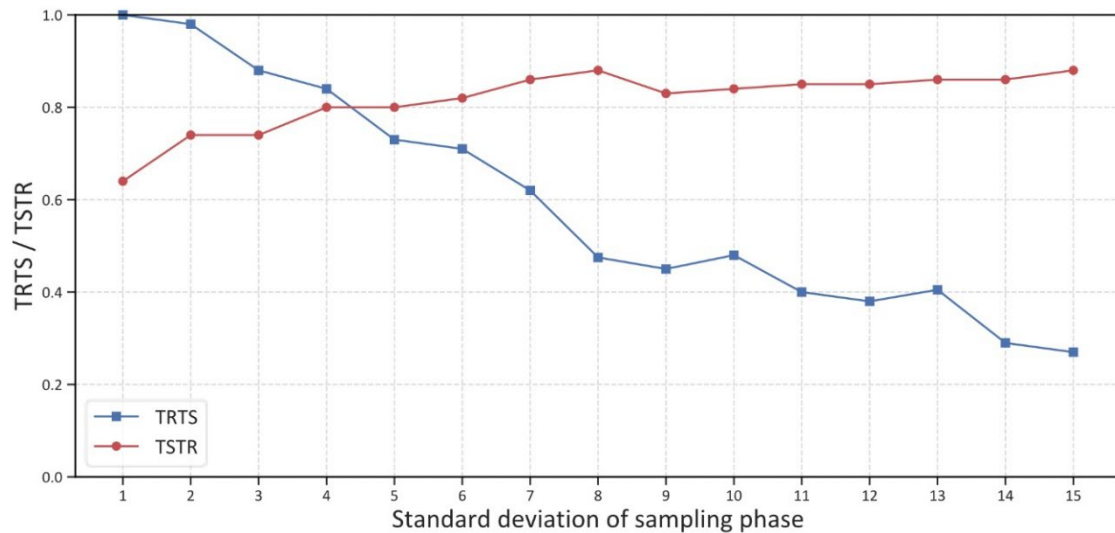


Figure 5: Relationship between standard deviation of the gaussian latent sampling and the Rocket classification metrics TRTS / TSTR (train on real, test on synthetic / train on synthetic, test on real)

According to these metrics, it is indeed feasible to use a low-complexity cVAE model architecture even for phase-averaged biometric data. This could potentially save large amounts of data within the framework of automated biometric authentication methods, thereby addressing the ever-increasing demand for computational power by AI models. For a more general result apart from this case-study more experiments with different classifier algorithms and larger datasets are recommended.

5. Conclusion and Outlook

This work investigates the feasibility of a low-complexity cVAE model for the generation of synthetic biometric gait analysis data. The model is provided only with reduced data, achieved through phase-averaging. The experimental results indicate that it is still possible to extract enough distinct features from the phase-averaged data for classification through state-of-the-art classification models. To confirm the general feasibility of containing biometric information in the phase-averaged synthetic data larger-scale experiments will be necessary.

Future work might tackle more computationally complex methods like diffusion models and generative adversarial networks.

Acknowledgement

This work has been supported by the Federal Ministry of Education and Research of the Federal Republic of Germany Förderkennzeichen 16KIS2239K, Sustained). The authors alone are responsible for the content of the paper.

References

- Agrawal, G., Kaur, A. and Myneni, S. (2024) 'A Review of Generative Models in Generating Synthetic Attack Data for Cybersecurity', *Electronics*, 13(2), p. 322. Available at: <https://doi.org/10.3390/electronics13020322>.
- Dempster, A., Petitjean, F. and Webb, G.I. (2020) 'ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels', *Data Mining and Knowledge Discovery*, 34(5), pp. 1454–1495. Available at: <https://doi.org/10.1007/s10618-020-00701-z>.
- Esteban, C., Hyland, S.L. and Rätsch, G. (2017) 'Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs'. Available at: <https://doi.org/10.48550/ARXIV.1706.02633>.
- Feiste, P. (2004) *Vergleichende Analyse des Gangbildes bei Patienten mit degenerativer Gonarthrose vor und nach Implantation einer Kniegelenkendoprothese mit Hilfe der Ganganalyse*. Greifswald university.
- Fortuin, V. et al. (2020) 'GP-VAE: Deep Probabilistic Time Series Imputation', in: *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, PMLR 108:1651-1661*.
- Goodfellow, I. et al. (2014) 'Generative Adversarial Nets', in Z. Ghahramani et al. (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

- Herbst, J. *et al.* (2024) 'One Step Towards Secure Identification in Wireless Body Area Networks (WBANs): Intelligent Insole Sensor Systems', in *2024 IEEE International Conference on Flexible and Printable Sensors and Systems (FLEPS)*. *2024 IEEE International Conference on Flexible and Printable Sensors and Systems (FLEPS)*, Tampere, Finland: IEEE, pp. 1–4. Available at: <https://doi.org/10.1109/FLEPS61194.2024.10603948>.
- Heusel, M. *et al.* (2017) 'GANs trained by a two time-scale update rule converge to a local nash equilibrium', in *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc. (NIPS'17), pp. 6629–6640.
- Kingma, D.P. and Welling, M. (2014) 'Auto-Encoding Variational Bayes'. ICLR 2014. Available at: <https://doi.org/10.48550/arXiv.1312.6114>.
- Kong, Z. *et al.* (2021) 'DiffWave: A Versatile Diffusion Model for Audio Synthesis', in *International Conference on Learning Representations (ICLR)*.
- Lee, D., Malacarne, S. and Aune, E. (2023) 'Vector Quantized Time Series Generation with a Bidirectional Prior Model', in F. Ruiz, J. Dy, and J.-W. van de Meent (eds) *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*. PMLR (Proceedings of Machine Learning Research), pp. 7665–7693. Available at: <https://proceedings.mlr.press/v206/lee23d.html>.
- Middlehurst, M. *et al.* (2021) 'HIVE-COTE 2.0: a new meta ensemble for time series classification', *Machine Learning*, 110(11–12), pp. 3211–3243. Available at: <https://doi.org/10.1007/s10994-021-06057-9>.
- Osadchy, M. *et al.* (2017) 'GenFace: Improving Cyber Security Using Realistic Synthetic Face Generation', in S. Dolev and S. Lodha (eds) *Cyber Security Cryptography and Machine Learning*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 19–33. Available at: https://doi.org/10.1007/978-3-319-60080-2_2.
- Papavasileiou, I. *et al.* (2021) 'GaitCode: Gait-based continuous authentication using multimodal learning and wearable sensors', *Smart Health*, 19, p. 100162. Available at: <https://doi.org/10.1016/j.smhl.2020.100162>.
- Reddy, G.T. *et al.* (2020) 'Analysis of Dimensionality Reduction Techniques on Big Data', *IEEE Access*, 8, pp. 54776–54788. Available at: <https://doi.org/10.1109/ACCESS.2020.2980942>.
- Rüb, M. *et al.* (2022) 'No One Acts like You: AI based Behavioral Biometric Identification', in *2022 3rd International Conference on Next Generation Computing Applications (NextComp)*. *2022 3rd International Conference on Next Generation Computing Applications (NextComp)*, Flic-en-Flac, Mauritius: IEEE, pp. 1–7. Available at: <https://doi.org/10.1109/NextComp55567.2022.9932247>.
- Singh, A., Prakash, B.S. and Chandrasekaran, K. (2016) 'A comparison of linear discriminant analysis and ridge classifier on Twitter data', in *2016 International Conference on Computing, Communication and Automation (ICCCA)*. *2016 International Conference on Computing, Communication and Automation (ICCCA)*, Greater Noida, India: IEEE, pp. 133–138. Available at: <https://doi.org/10.1109/CCAA.2016.7813704>.
- Sohn, K., Lee, H. and Yan, X. (2015) 'Learning Structured Output Representation using Deep Conditional Generative Models', in C. Cortes *et al.* (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper_files/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf.
- Sowmya, T. and Mary Anita, E.A. (2023) 'A comprehensive review of AI based intrusion detection system', *Measurement: Sensors*, 28, p. 100827. Available at: <https://doi.org/10.1016/j.measen.2023.100827>.
- 'Stappone' (2025) in. <https://www.stappone.com/produkte/digitale-ganganalyse-orthopaedie-neurologie/stappone-research/> (Accessed: 09.01.2025).