LITHUANIAN COMPUTER SOCIETY
VILNIUS UNIVERSITY, INSTITUTE OF DATA SCIENCE AND DIGITAL TECHNOLOGIES
LITHUANIAN ACADEMY OF SCIENCES

16th Conference on

DATA ANALYSIS METHODS
for Software Systems

November 27–29, 2025

Druskininkai, Lithuania, Hotel "Europa Royale"
https://www.mii.lt/DAMSS

VILNIUS UNIVERSITY PRESS
Vilnius, 2025

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernataviciene@mif.vu.lt
Prof. Olga Kurasova
olga.kurasova@mif.vu.lt
Tel. (+370 5) 2109 315

https://doi.org/10.15388/DAMSS.16.2025
ISBN 978-609-07-1200-9 (digital PDF)

Comparative Analysis of ECG Data Augmentation Methods in Arrhythmia Classification

Jonas Mindaugas Rimšelis, Povilas Treigys, Zigmantas Kęstutis Juškevičius, Jolita Bernatavičienė

Institute of Data Science and Digital Technologies
Vilnius University

jonas.rimselis@mif.vu.lt

An electrocardiogram (ECG) measures electrical signals from the heart to capture various cardiovascular conditions. Distinct patterns arise in ECG during abnormal heartbeats, which facilitate the recognition of cardiovascular diseases through non-invasive ECG. Single-lead Holter devices allow uninterrupted, continuous monitoring of heart performance during everyday tasks and the identification of cardiovascular diseases. Deep learning methods are utilized for classifying heartbeats and raising awareness of deteriorating health [1]. Since abnormal heartbeats occur rarely, even in patients diagnosed with arrhythmia, data used for training models are imbalanced, leading to poor generalization and robustness [2]. Data augmentation is utilized to mitigate label balancing issues. Data augmentation techniques can be divided into traditional augmentation and generative deep learning methods. While traditional augmentation utilizes transformations of existing data to synthesize training data, generative methods utilize Generative Adversarial Networks (GANs) [3], Variational Autoencoders (VAEs) [4], and Diffusion Discrete Probabilistic Models (DDPMs) to create artificial signals [5]. As a traditional augmentation technique, SMOTE has been applied to ECG datasets, but some practitioners have raised concerns that it may implicitly distort morphological or temporal properties of ECG signals due to its interpolation mechanism. In contrast, generative methods tend to synthesize signals that mimic real-world data but tend to simplify signal morphology [6]. Furthermore, there is a lack of research on synergies between preprocessing and data augmentation techniques. In this study, a literature review is performed to capture the most prominent

and efficient data augmentation methods for ECG considering heartbeat classification in arrhythmia cases. Furthermore, synergies between preprocessing and data augmentation methods are analyzed. The review is followed by a comparative analysis of leading augmentation approaches, focusing particularly on ECG signals generated using DDPMs for the MIT-BIH Arrhythmia Database classification task. It is hypothesized that a 1Ds Convolutional Neural Network (CNN) classifier will show better performance in abnormal beat classification when trained on data augmented by DDPM than by other methods.

Acknowledgements. Funding was provided by the Research Council of Lithuania (LMTLT), agreement Nr. S-ITP-25-9.

References:

[1] O. A. Oke and N. Cavus, "A systematic review on the impact of artificial intelligence on electrocardiograms in cardiology," Int. J. Med. Inform., vol. 195, p. 105753, 2025, doi: https://doi.org/10.1016/j.ijmedinf.2024.105753.

[2] Z. Chen et al., "A novel imbalanced dataset mitigation method and ECG classification model based on combined 1D_CBAM-autoencoder and lightweight CNN model," Biomed. Signal Process. Control, vol. 87, no. PB, p. 105437, 2024, doi: 10.1016/j.bspc.2023.105437.

[3] M. Kuntalp and O. Düzyel, "A new method for GAN-based data augmentation for classes with distinct clusters," Expert Syst. Appl., vol. 235, no. March 2023, p. 121199, 2024, doi: 10.1016/j.eswa.2023.121199.

[4] Y. Xia, W. Wang, and K. Wang, "ECG signal generation based on conditional generative models," Biomed. Signal Process. Control, vol. 82, no. May 2022, p. 104587, 2023, doi: 10.1016/j.bspc.2023.104587.

[5] E. Adib, A. S. Fernandez, F. Afghah, and J. J. Prevost, "Synthetic ECG Signal Generation Using Probabilistic Diffusion Models," IEEE Access, vol. 11, no. May, pp. 75818–75828, 2023, doi: 10.1109/ACCESS.2023.3296542.

[6] L. Simone, D. Bacciu, and V. Gervasi, "ECG synthesis for cardiac arrhythmias: Integrating self-supervised learning and generative adversarial networks," Artif. Intell. Med., vol. 167, no. May 2024, p. 103162, 2025, doi: 10.1016/j.artmed.2025.103162.