EEG Biometrics with GAN Integration for Secure Smart City Data Access

Roberto Saia, Riccardo Balia, Alessandro Sebastian Podda, Livio Pompianu, Salvatore Carta, Alessia Pisu
Department of Mathematics and Computer Science, University of Cagliari, Italy
Artificial Intelligence and Big Data Laboratory (*https://aibd.unica.it*)
{roberto.saia,riccardo.balia,sebastianpodda,livio.pompianu,salvatore,alessia.pisu96}@unica.it

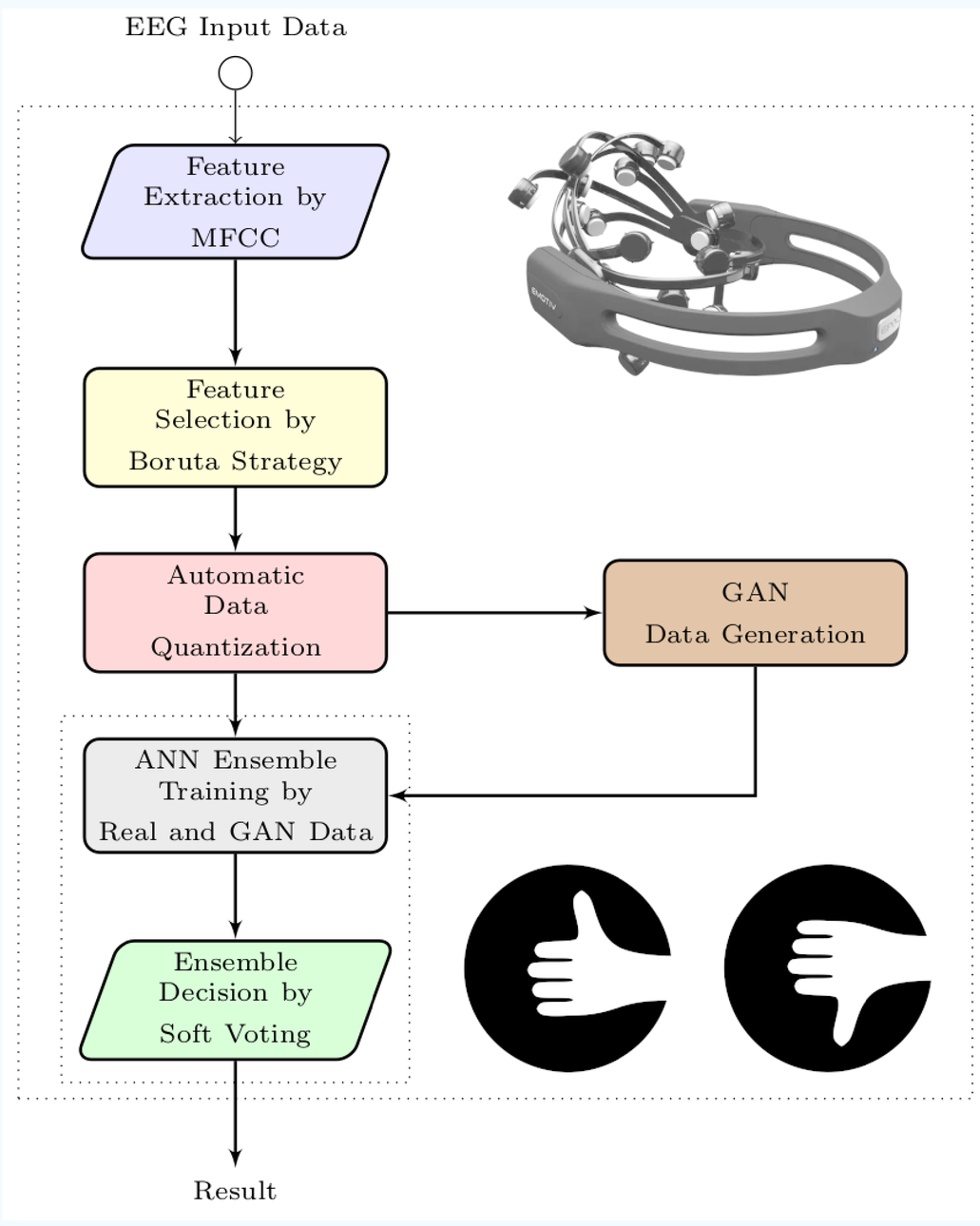
Abstract

Biometric systems using **EEG** data for user authentication show promise, particularly in **Smart City applications** where secure access to sensitive data is vital (e.g., healthcare, transportation, and public safety). However, EEG data complexity poses **challenges**. This work introduces a novel user verification framework, building on preliminary experiments with the Biometric EEG Dataset (BED). It employs **Mel-Frequency Cepstral Coefficients** (MFCC) for feature extraction, followed by the **Boruta** method for feature selection and **data quantization**. A key innovation is the use of a **Generative Adversarial Network** (GAN) to generate synthetic EEG data, which, combined with real data, trains an ensemble of **Artificial Neural Network**s (ANNs). The system makes decisions through **soft voting**, aiming to provide a robust and competitive solution. Initial results show **strong potential** for further refinement and improvement.

Research Scenario

- The proposed approach would like **to address the limitation** of existing methods, which often lack practicality for large-scale use due to their **complexity**, lengthy **acquisition times**, and a **small number of participants**.
- By focusing on **low cost**, **simplicity**, and **efficiency**, it aims to offer a solution suitable for **large-scale** deployment, while ensuring validation against state-of-the-art techniques.

Approach Architecture



Approach Definition

- **EEG Input Data**: The adopted Biometric EEG Dataset (BED) contains EEG signals collected with the *Emotiv Epoc+* wireless headset. The dataset includes three feature extraction methods (MFCC, ARRC, SPEC) and various stimuli such as visual images, cognitive tasks, and visual evoked potentials across different frequencies.
- **Feature Extraction**: After preliminary experimentation with several methods, we select *MFCC* as the most promising feature extraction technique due to its potential for performance improvements when combined with our proposed approach.
- **Feature Selection**: We use the Boruta technique to detect the most relevant features by comparing their importance to shadow ones, ensuring only the most informative data is retained and reducing overfitting.
- **Automatic Data Quantization**: This step automatically determines the optimal number of bins to transform continuous features into discrete categories, simplifying the feature space and enhancing performance through K-fold cross-validation.
- **GAN Data Generation**: A Generative Adversarial Network (GAN) generates synthetic EEG data that mimics real samples, expanding the training dataset and improving model generalization.
- **ANN Ensemble Training)**: Artificial Neural Networks (ANNs) are trained on both real and GAN-generated synthetic data using K-fold cross-validation. An early stopping mechanism ensures model robustness, and predictions are refined using soft voting.
- **Ensemble Decision**: The ensemble model uses soft voting, considering the confidence levels of individual classifiers, resulting in a more nuanced and probabilistic decision-making process.

Conclusions

- **EEG-Based Authentication**: We propose an EEG-based biometric system for user authentication, aimed at secure access to sensitive information in Smart City ecosystems. The system identifies individuals by their brain activity patterns, offering a reliable and robust solution.
- **GAN-Enhanced Data Augmentation**: Generative Adversarial Networks (GANs) are used to generate synthetic EEG data, addressing data variability and scarcity issues, thus improving the model's robustness and generalizability.
- **Ensemble ANN Classification**: The architecture includes feature extraction (MFCC), feature selection (Boruta), and data quantization. An ensemble of Artificial Neural Networks (ANNs), trained on real and synthetic data, enhances prediction accuracy using soft voting.

Acknowledgements

We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 - Call for tender No.3277 published on December 30, 2021 by the Italian Ministry of University and Research (MUR) funded by the European Union – NextGenerationEU. Project Code ECS0000038 – Project Title eINS Ecosystem of Innovation for Next Generation Sardinia – CUP F53C22000430001 - Grant Assignment Decree No. 1056 adopted on June 23, 2022 by the Italian Ministry of Ministry of University and Research (MUR).