



**Politechnika  
Śląska**

**Wydział Automatyki, Elektroniki  
i Informatyki**

Praca Magisterska

Plan działania

Zastosowanie metod sztucznej inteligencji do detekcji  
arytmii na podstawie sygnałów PPG

Jakub Kula

Gliwice 2025

Przygotowałem krótki dokument, w którym przedstawiam swoje założenia dotyczące części projektowo-badawczej, aby upewnić się, że są one zgodne z Pańskimi oczekiwaniami.

## 1 Wybrane bazy danych

- MIMIC PERform AF Dataset
- Multiclass Arrhythmia Detection and Classification(...)
- Dane wygenerowane dzięki Symulatorzes sygnału PPG z epizodami arrhythmia

## 2 Wybrane cechy

Dla celów badawczych zostały wybrano 30 cech. Niektóre z nich mogą okazać się nadmiarowe, więc praca w kolejnych krokach będzie zmniejszała ich ilość wybierając te najważniejsze przy pomocy funkcji "SelectKBest" z sklearn lub/i innych metod wyjaśnialnej sztucznej inteligencji. Sprawdzane opcje to 30-25-20-15-10.

Nie jestem pewny poprawności nazw cech - opracowywałem je z przy pomocy List of features extracted from the PPG signal that is used for AF detection.

- 1. Średnia

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

Gdzie: -  $x_i$  - wartość  $i$ -tej próbki sygnału, -  $N$  - liczba próbek.

- 2. Mediana
- 3. Odchylenie standardowe ( $\sigma$ )

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

- 4. Wariancja (var)

$$\text{var} = \sigma^2$$

- 5. Rozstęp międzykwartyłowy (iqr)

$$\text{iqr} = Q_3 - Q_1$$

Gdzie: -  $Q_1$  - pierwszy kwartył, -  $Q_3$  - trzeci kwartył.

- 6. Maksimum
- 7. Minimum
- 8. Średnia wartość modułu różnic

$$\overline{|\text{diff}|} = \frac{1}{N-1} \sum_{i=1}^{N-1} |x_{i+1} - x_i|$$

- **9. Procent różnic dodatnich**

$$\frac{100 \sum_{i=1}^{N-1} \mathbb{I}((x_{i+1} - x_i) > 0)}{N - 1}$$

Gdzie: -  $\mathbb{I}(\cdot)$  – funkcja indykatorowa (1 jeśli warunek jest spełniony, 0 w przeciwnym przypadku).

- **10. RMS różnic (RMSSD)**

$$\text{RMSSD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}$$

- **11. Średnie odchylenie absolutne**

$$\text{mad} = \frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}|$$

- **12. Energia**

$$\sum_{i=1}^N x_i^2$$

- **13. Skośność (skewness)**

$$\text{skewness} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{\sigma^3}$$

- **14. Entropia Shannona (wentropy)**

$$\text{wentropy} = - \sum_{i=1}^N p(x_i) \log p(x_i)$$

Gdzie: -  $p(x_i)$  – prawdopodobieństwo wystąpienia wartości  $x_i$ .

- **15. Szczyt widmowy**

$$\text{maximal spectral peak} = \arg \max_f P_{xx}(f)$$

Gdzie: -  $P_{xx}(f)$  – moc widmowa dla częstotliwości  $f$ .

- **16. Średnia mocy widmowej**

$$\overline{P_{xx}} = \frac{1}{M} \sum_{j=1}^M P_{xx}(f_j)$$

Gdzie: -  $M$  – liczba częstotliwości w widmie.

- **17. Odchylenie standardowe widma**

$$\sigma(P_{xx}) = \sqrt{\frac{1}{M} \sum_{j=1}^M (P_{xx}(f_j) - \overline{P_{xx}})^2}$$

- 18. Kurtoza widma

$$\text{kurtosis}(P_{xx}) = \frac{\frac{1}{M} \sum_{j=1}^M (P_{xx}(f_j) - \overline{P_{xx}})^4}{\sigma(P_{xx})^4}$$

- 19. Całkowita energia widmowa

$$\sum_{j=1}^M P_{xx}(f_j)$$

- 20. Udział wysokich pików

$$\frac{\sum_{j=1}^M \mathbb{I}(P_{xx}(f_j) > \overline{P_{xx}})}{M}$$

- 21. Średnia wartość współczynnika falkowego

$$\overline{|wt|} = \frac{1}{K} \sum_{k=1}^K |wt_k|$$

Gdzie: -  $wt_k$  – współczynnik falkowy dla  $k$ -tej skali.

- 22. Odchylenie standardowe współczynnika falkowego

$$\sigma(|wt|) = \sqrt{\frac{1}{K} \sum_{k=1}^K (|wt_k| - \overline{|wt|})^2}$$

- 23. Energia falkowa

$$\sum_{k=1}^K |wt_k|^2$$

- 24. Współczynnik zmienności

$$\frac{\overline{x}}{\sigma}$$

- 25. Maksymalny współczynnik falkowy

- 26. Mediana współczynnika falkowego

- 27. Entropia Shannona współczynników falkowych

$$\text{wentropy}(|wt|) = - \sum_{k=1}^K p(|wt_k|) \log p(|wt_k|)$$

- 28. Znormalizowana średnia wartość modułu różnic

$$\frac{\frac{1}{N} \sum_{i=1}^{N-1} |x_{i+1} - x_i|}{\overline{x}}$$

- 29. Znormalizowana suma różnic

$$\frac{\sum_{i=1}^{N-1} |x_{i+1} - x_i|}{\sum_{i=1}^N |x_i|}$$

- 30. Znormalizowany RMSSD

$$\frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}}{\sqrt{\overline{x^2}}}$$

## 2.1 Kod wyciągający cechy

```
1 featureMatrix(i, 1) = mean(signal); % Mean
2 featureMatrix(i, 2) = median(signal); % Median
3 featureMatrix(i, 3) = std(signal); % Standard Deviation
4 featureMatrix(i, 4) = var(signal); % Variance
5 featureMatrix(i, 5) = iqr(signal); % Interquartile Range
6 featureMatrix(i, 6) = max(signal); % Maximum
7 featureMatrix(i, 7) = min(signal); % Minimum
8 featureMatrix(i, 8) = mean(abs(diff(signal))); % Average of
   the absolute value of the differences
9 featureMatrix(i, 9) = 100 * sum(diff(signal) > 0) / (length(
   signal) - 1); % Percentage of Positive Differences
10 featureMatrix(i, 10) = sqrt(mean(diff(signal).^2)); % RMSSD
11 featureMatrix(i, 11) = mad(signal, 1); % Mean Absolute
   Deviation
12 featureMatrix(i, 12) = sum(signal.^2); % Energy
13 featureMatrix(i, 13) = skewness(signal); % Asymmetry
14
15 % --- Entropia ---
16 featureMatrix(i, 14) = wentropy(signal, 'shannon'); % Shannon
   Entropy
17
18 % --- Cechy z domeny częstotliwości ---
19 [pxx, f] = periodogram(signal); % Widmo sygnału
20 featureMatrix(i, 15) = f(pxx == max(pxx)); % Maximal Spectral
   Peak
21 featureMatrix(i, 16) = mean(pxx); % Mean of Spectrum
22 featureMatrix(i, 17) = std(pxx); % Standard Deviation of
   Spectrum
23 featureMatrix(i, 18) = kurtosis(pxx); % Kurtosis of Spectrum
24 featureMatrix(i, 19) = sum(pxx); % Total Spectral Energy
25 featureMatrix(i, 20) = sum(pxx > mean(pxx)) / length(pxx); %
   Fraction of High Peaks
26
27 % --- Zależności czasowo-częstotliwościowe ---
28 wt = cwt(signal, 'amor'); % Continuous Wavelet Transform (
   Morlet Wavelet)
29 featureMatrix(i, 21) = mean(abs(wt), 'all'); % Mean Wavelet
   Coefficient Magnitude
30 featureMatrix(i, 22) = std(abs(wt), 0, 'all'); % Std Wavelet
   Coefficient Magnitude
31 featureMatrix(i, 23) = sum(abs(wt).^2, 'all'); % Wavelet
   Energy
32
33 % --- Cechy dodatkowe ---
34 featureMatrix(i, 24) = featureMatrix(i, 1) / featureMatrix(i,
   3); % Coefficient of Variation
35 featureMatrix(i, 25) = max(abs(wt), [], 'all'); % Max Wavelet
   Coefficient
```

```

36 featureMatrix(i, 26) = median(abs(wt), 'all'); % Median
    Wavelet Coefficient
37 featureMatrix(i, 27) = wentropy(abs(wt), 'shannon'); %
    Wavelet Shannon Entropy
38
39 % Normalizacja cech związanych z różnicami
40 differences = diff(signal);
41 featureMatrix(i, 28) = mean(abs(differences)) / featureMatrix
    (i, 1); % Normalized Absolute Deviation
42 featureMatrix(i, 29) = sum(abs(differences)) / sum(abs(signal
    )); % Normalized Absolute Difference
43 featureMatrix(i, 30) = sqrt(mean(differences.^2)) / sqrt(mean
    (signal.^2)); % Normalized RMSSD

```

### 3 Techniki uczenia maszynowego

Do celów klasyfikacji wybrałem 5 metod klasycznych oraz po kilka hiperparametrów które będą dostrajane. Dostrajanie będzie wykonywane przy pomocy zbioru walidującego, który będzie stanowić 15%-20% zbioru uczącego - tak aby nie doszło do przecieku danych

#### 3.1 Drzewa decyzyjne

- `max_depth` – maksymalna głębokość drzewa
- `min_samples_split` – minimalna liczba próbek do podziału węzła
- `min_samples_leaf` – minimalna liczba próbek w liściu drzewa
- `max_features` – liczba cech do rozważenia przy podziale węzła

#### 3.2 Random Forest

- `n_estimators` – liczba drzew w lesie
- `max_depth` – maksymalna głębokość każdego drzewa
- `min_samples_split` – minimalna liczba próbek do podziału węzła
- `min_samples_leaf` – minimalna liczba próbek w liściu drzewa
- `max_features` – liczba cech rozważanych przy podziale węzła

#### 3.3 K-Nearest Neighbors

- `n_neighbors` – liczba sąsiadów do rozważenia przy klasyfikacji
- `metric` – miara odległości

### 3.4 Support Vector Machines

- `C` – parametr regularyzacji
- `kernel` – rodzaj jądra
- `degree` – stopień wielomianu
- `gamma` – współczynnik jądra
- `coef0` – współczynnik dla jądra wielomianowego i sigmoidalnego
- `probability` – czy obliczać prawdopodobieństwa
- `shrinking` – czy używać algorytmu kurczenia

### 3.5 Naïve Bayes

### 3.6 Sieci neuronowe

Sieci neruonowe będą dostrajane przy użyciu biblioteki "Optuna" która pozwala stworzyć nam eksperyment którego celem jest maksymalizacja parametru zadanego. W naszym przykładzie będzie to maksymalizacja dokładności albo F1 dla zbioru walidującego

Wybrane architektury:

- Gęste sieci neuronowe
- Konwolucyjne sieci neuronowe
- Rekurencyjne sieci neuronowe
- LSTM

## 4 Podsumowanie

Ze względu na początek semestru oraz związane z nim kwestie organizacyjne, rozpocznę prace nad częścią projektowo-badawczą za kilka dni. Pozwoli mi to na spokojne opracowanie harmonogramu, który uwzględni obowiązki akademickie, pracę nad pracą dyplomową oraz działalność w kole naukowym.

Jeśli ma Pan jakiegokolwiek uwagi, pytania lub sugestie dotyczące moich założeń, chętnie umówię się na spotkanie lub rozmowę przez Microsoft Teams w dogodnym dla Pana terminie.