

Klasyfikacja sygnałów EKG ze zbioru PTB-XL przy użyciu metod Głębokiego Uczzenia

Mateusz Stasiak, Daminika Dzeranhouskaya

Styczeń 2026

Spis treści

1 Wstęp	2
1.1 Charakterystyka zbioru danych PTB-XL	2
1.1.1 Przygotowanie etykiet i definicje klas	2
1.1.2 Strategia podziału danych)	3
1.1.3 Analiza eksploracyjna i charakterystyka demograficzna	4
1.1.4 Wizualna analiza sygnałów i motywacja dla metod Deep Learning .	4
2 Opis teoretyczny rozwiązania	6
2.1 Model Bazowy: Random Forest (Baseline)	6
2.2 Architektury Głębokiego Uczzenia	6
2.3 Transformacja danych i kanały pochodne	7
2.4 Strategia Augmentacji Danych	7
3 Eksperymenty i analiza wyników	8
3.1 Eksperyment 1: Baseline vs Deep Learning	8
3.2 Eksperyment 2: Porównanie architektur ResNet i InceptionTime	8
3.3 Eksperyment 3: Wpływ augmentacji i pochodnych sygnału	9
3.4 Analiza jakości predykcji na podstawie macierzy pomyłek	10
4 Interpretowalność modelu – Saliency Maps	11
4.1 Przypadek 1: Prawidłowa detekcja zawału (Klasa MI)	12
4.2 Przypadek 2: Błędna klasyfikacja zaburzeń przewodzenia (Klasa CD)	13
5 Podsumowanie i wnioski	14

1 Wstęp

Elektrokardiografia (EKG) stanowi podstawowe narzędzie diagnostyczne w kardiologii, pozwalające na nieinwazyjną ocenę aktywności elektrycznej serca. Wraz z rosnącą objętością danych medycznych, automatyzacja analizy zapisów EKG stała się kluczowym wyzwaniem z pogranicza medycyny i informatyki.

Celem niniejszego projektu jest opracowanie i porównanie algorytmów uczenia maszynowego służących do automatycznej klasyfikacji wielokanałowych sygnałów EKG. Praca opiera się na zbiorze danych **PTB-XL**, zawierającym kliniczne zapisy od ponad 18 000 pacjentów.

W ramach projektu przeprowadzono szereg eksperymentów obejmujących:

- Opracowanie modelu bazowego (*Baseline*) opartego na klasycznych metodach uczenia maszynowego (Random Forest) oraz ekstrakcji cech statystycznych.
- Implementację zaawansowanych architektur głębokich sieci neuronowych: **ResNet1D** oraz **InceptionTime**.
- Analizę wpływu augmentacji danych oraz wykorzystania pochodnych sygnału na jakość klasyfikacji.
- Weryfikację interpretowalności modeli przy użyciu map istotności (Saliency Maps), co jest kluczowe w zastosowaniach medycznych.

Niniejszy raport dokumentuje proces konstrukcji modeli, uzyskane wyniki metryczne oraz analizę statystyczną istotności różnic między zaproponowanymi rozwiązaniami.

1.1 Charakterystyka zbioru danych PTB-XL

W projekcie wykorzystano zbiór **PTB-XL**, będący jednym z największych publicznie dostępnych zbiorów klinicznych sygnałów EKG. Zawiera on 21 837 rekordów od 18 885 pacjentów. Zdecydowano się na pracę na danych próbkoowych z częstotliwością **100 Hz**, co pozwala na zachowanie kluczowych cech morfologicznych sygnału przy jednoczesnej optymalizacji czasu trenowania modeli głębokich.

1.1.1 Przygotowanie etykiet i definicje klas

Proces etykietowania wymagał agregacji szczegółowych kodów SCP-ECG do pięciu głównych kategorii diagnostycznych (*Superclasses*). Poniżej przedstawiono charakterystykę ostatecznie wybranych klas:

- **NORM (Normal ECG)**: Zapisy prawidłowe, charakteryzujące się właściwym rytem zatokowym oraz brakiem zmian morfologicznych w obrębie załamków i odcinków.
- **MI (Myocardial Infarction)**: Zawał mięśnia sercowego. Klasa obejmuje zarówno ostre stany z uniesieniem odcinka ST, jak i ślady po przebytych zawałach (np. patologiczne załamki Q).
- **STTC (ST/T Changes)**: Niespecyficzne zmiany w obrębie odcinka ST oraz załamka T. Mogą one sugerować niedokrwienie mięśnia sercowego lub zaburzenia elektrolitowe.

- **CD (Conduction Disturbance):** Zaburzenia przewodzenia, w tym bloki odnów pęczka Hisa (RBBB, LBBB) oraz inne nieprawidłowości w szerzeniu się impulsu elektrycznego w komorach.
- **HYP (Hypertrophy):** Cechy przerostu komór serca, objawiające się zazwyczaj zwiększoną amplitudą zespołów QRS w określonych odprowadzeniach.

1.1.2 Strategia podziału danych)

W celu zapewnienia najwyższej wiarygodności wyników oraz uniknięcia wycieku danych, zaimplementowano autorską procedurę podziału zbioru, która wykracza poza standar-dowy podział losowy. Procedura ta opiera się na dwóch kluczowych filarach: unikalności pacjentów oraz priorytetyzacji danych zwalidowanych przez ekspertów.

1. **Izolacja na poziomie pacjenta:** Podział danych jest realizowany w oparciu o identyfikator `patient_id`. Gwarantuje to, że żadne badanie należące do konkret-nego pacjenta nie pojawi się w więcej niż jednym zbiorze. Jest to kluczowe w analizie sygnałów EKG, gdyż model mógłby nauczyć się rozpoznawać unikalne cechy osobnicze pacjenta zamiast ogólnych cech patologii.
2. **Priorytetyzacja danych zwalidowanych przez lekarza:** Zbiór PTB-XL zawiera flagę `validated_by_human`. W opracowanym algorytmie dokonano segmentacji pa-cjentów na dwie grupy:
 - **Grupa A:** Pacjenci, których wszystkie badania zostały w 100% zwalidowane przez kardiologa.
 - **Grupa B:** Pozostali pacjenci, posiadający przynajmniej jedno badanie bez pełnej walidacji eksperckiej.
3. **Konstrukcja zbiorów Val i Test:** Aby zbiory służące do oceny modelu były najwyższej jakości, zbiory walidacyjny oraz testowy (każdy stanowiący ok. 15% ca-łości) są budowane **wyłącznie** z pacjentów należących do Grupy A. Gwarantuje to, że ostateczna ewaluacja modelu odbywa się na danych o najwyższym stopniu pewności diagnostycznej.
4. **Konstrukcja zbioru treningowego:** Zbiór treningowy (ok. 70% danych) składa się z pozostałych pacjentów z Grupy A oraz wszystkich pacjentów z Grupy B. Takie podejście pozwala modelowi uczyć się na dużym, zróżnicowanym wolumenie danych, zachowując jednocześnie rygorystyczne warunki dla fazy testowej.

Finalny rozkład danych przedstawia się następująco:

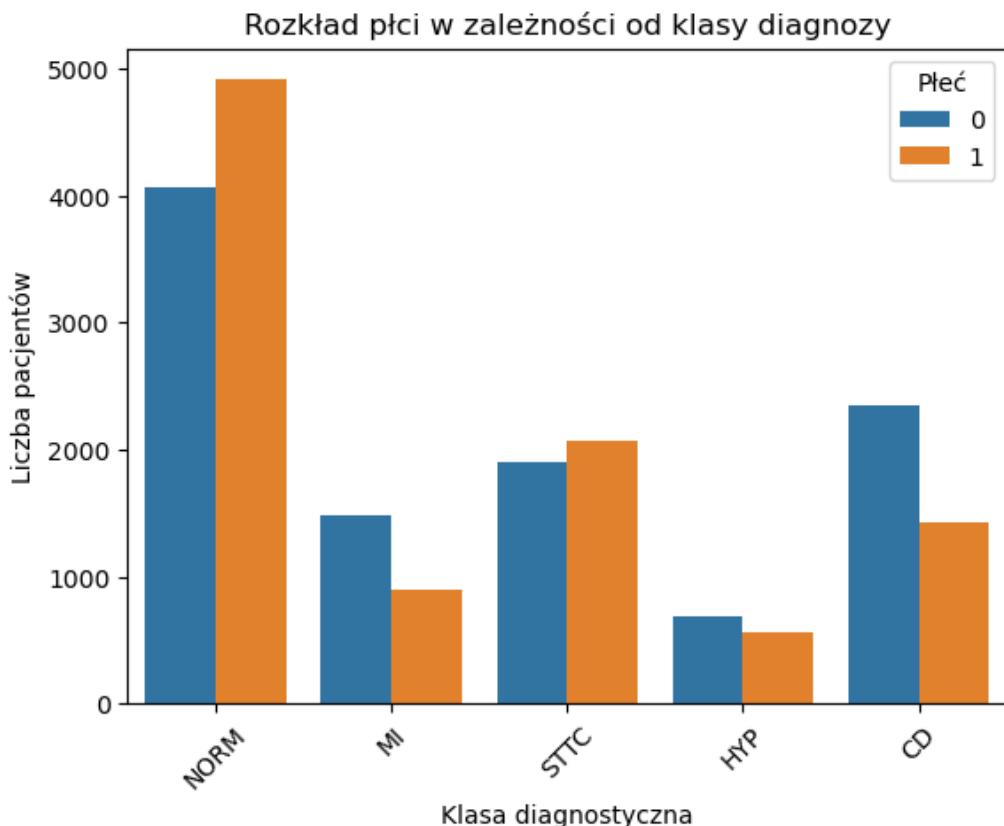
Tabela 1: Statystyki autorskiego podziału danych.

Zbiór	Liczba próbek	% Całości	% Walidacji ludzkiej
Treningowy (Train)	15 285	70,00%	≈ 42%
Walidacyjny (Val)	3 276	15,00%	100%
Testowy (Test)	3 276	15,00%	100%

Tak zaprojektowany potok przetwarzania danych zapewnia, że raportowane metryki jakości są rzetelnym odzwierciedleniem zdolności modelu do generalizacji wiedzy na nowych, klinicznie zweryfikowanych przypadkach.

1.1.3 Analiza eksploracyjna i charakterystyka demograficzna

Wstępna analiza eksploracyjna danych (EDA) ujawniła istotną nierównowagę klasową w zbiorze PTB-XL. Dominującą kategorię stanowią rekordy prawidłowe (**NORM**), które odpowiadają za ok. **44,2%** całego zbioru. Pozostałe klasy rozkładają się następująco: zmiany ST/T (**STTC**) – **24,1%**, zawały mięśnia sercowego (**MI**) – **20,2%**, zaburzenia przewodzenia (**CD**) – **22,4%** oraz przerosty (**HYP**) – **12,2%**. Nierównowaga ta bezpośrednio wpływała na dobór metryk ewaluacyjnych — priorytetyzację miar **F1-Macro** oraz **Precision-Recall AUC**.



Rysunek 1: Rozkład płci w ramach poszczególnych klas diagnostycznych – 1 oznacza kobietę.

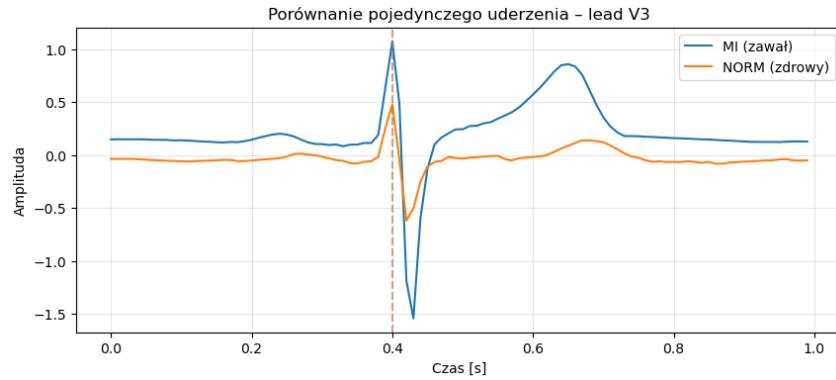
Analiza powyższego wykresu oraz korelacji cech demograficznych pozwoliła na sformułowanie następujących wniosków klinicznych:

- **Dymorfizm płciowy w patologiach:** Zaobserwowano wyraźną przewagę mężczyzn w klasach **MI** (zawał mięśnia sercowego) oraz **CD** (zaburzenia przewodzenia).
- **Rozkład w grupie STTC:** Notujemy że w ta klasa charakteryzuje się najbardziej zbalansowanym rozkładem płci.

1.1.4 Wizualna analiza sygnałów i motywacja dla metod Deep Learning

W ramach analizy EDA przeprowadzono wieloaspektowe porównanie sygnałów EKG dla różnych klas diagnostycznych. Celem było wykazanie, że surowe dane zawierają wyraźne różnice morfologiczne i statystyczne, co stanowi bezpośrednią motywację do zastosowania modeli głębokiego uczenia.

- **Analiza morfologii pojedynczego uderzenia:** Porównanie uśrednionego pojedynczego cyklu pracy serca (Rysunek 2) wykazuje znaczące różnice między klasą **NORM** a **MI**. W przypadku zawału mięśnia sercowego (*MI*) obserwuje się znacznie większe wychylenia sygnału oraz wyższą energię w obrębie zespołu QRS i odcinka ST. Te zmiany w amplitudzie i dynamice są kluczowymi markerami, które sieć CNN może identyfikować jako cechy dyskryminacyjne.

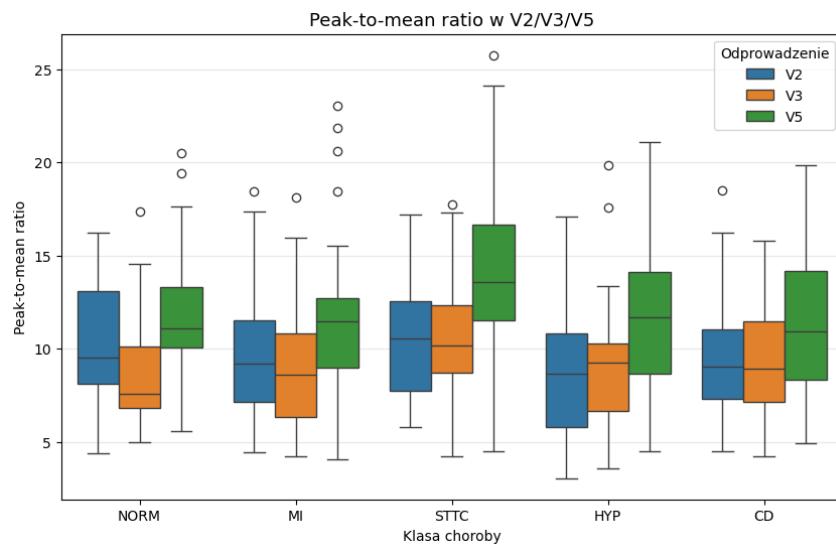


Rysunek 2: Porównanie morfologii pojedynczego uderzenia serca: klasa NORM vs MI.

- **Separacja klas poprzez Peak-to-Mean Ratio (PMR):** Zastosowano miarę *Peak-to-Mean Ratio*, która definiowana jest jako stosunek amplitudy szczytowej sygnału do jego wartości średniej (energii średniej):

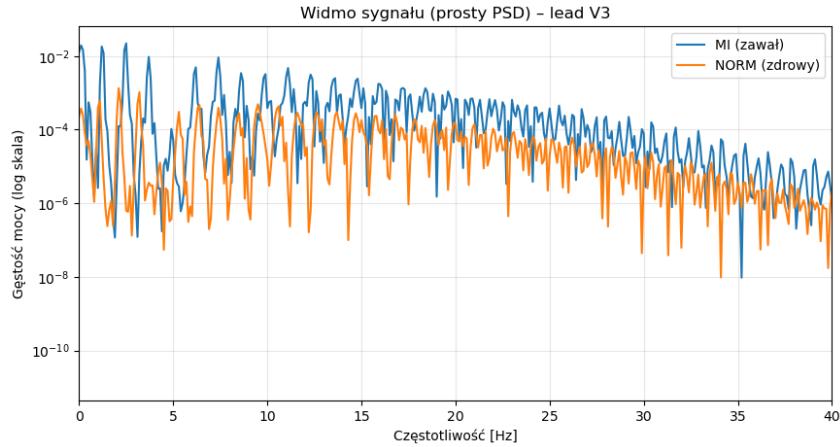
$$PMR = \frac{|x_{max}|}{\frac{1}{N} \sum_{i=1}^N |x_i|}$$

Jak przedstawiono na wykresie 3, wskaźnik ten pozwala na skutecną wizualną separację klas. Różne schorzenia zmieniają "ostrość" i "pikowość" zapisu EKG, co sprawia, że PMR staje się wartościową cechą statystyczną odróżniającą np. zdrowe serce od zaburzeń przewodzenia.



Rysunek 3: Rozkład wskaźnika Peak-to-Mean Ratio dla różnych klas diagnostycznych.

- **Analiza w domenie częstotliwości (FFT):** Wykorzystanie szybkiej transformaty Fouriera (Rysunek 4) pozwoliło na ocenę gęstości widmowej mocy sygnałów. Zaobserwowano, że sygnały różnią się energią na niemal wszystkich częstotliwościach w zależności od diagnozy. Przykładowo, klasy patologiczne wykazują inne składowe harmoniczne niż rytm zatokowy, co potwierdza, że informacja diagnostyczna jest rozproszona w całym spektrum sygnału.



Rysunek 4: Porównanie widma częstotliwościowego (FFT) dla wybranych kategorii diagnostycznych.

Powyższe obserwacje dowodzą, że różnice między klasami są zakodowane zarówno w czasie, amplitudzie, jak i częstotliwości. Sieci splotowe (CNN), dzięki warstwom filtrów o różnych rozmiarach, są w stanie automatycznie integrować te wszystkie aspekty, co uzasadnia ich wybór jako głównego narzędzia w niniejszym projekcie.

2 Opis teoretyczny rozwiązania

Wybór algorytmów został podyktowany chęcią zestawienia klasycznego podejścia opartego na inżynierii cech (*Feature Engineering*) z nowoczesnymi modelami głębokiego uczenia (*Deep Learning*), które samodzielnie ekstrahują istotne wzorce z surowego sygnału.

2.1 Model Bazowy: Random Forest (Baseline)

Jako punkt odniesienia przyjęto model lasu losowego (*Random Forest*). Operuje on na wektorze 12 cech statystycznych wyznaczonych dla każdego z 12 odprowadzeń (łącznie 144 cechy). Wybór ten uzasadnia wysoka odporność modelu na przeuczenie oraz zdolność do radzenia sobie z nieliniowymi zależnościami bez potrzeby rygorystycznej normalizacji danych.

2.2 Architektury Głębokiego Uczenia

Główna część projektu opiera się na dwóch architekturach splotowych 1D, które stanowią *State-of-the-Art* w analizie serii czasowych:

- **ResNet1D:** Wykorzystuje bloki rezydualne (*residual blocks*) z połączeniami skrótnymi (*skip connections*). Rozwiązuje to problem zanikającego gradientu, umożliwiając trenowanie głębokich sieci zdolnych do wychwytywania złożonych zależności morfologicznych w zapisie EKG.
- **InceptionTime:** Architektura ta wykorzystuje moduły Inception, które stosują równoległe sploty o różnych rozmiarach filtrów w jednej warstwie. Pozwala to modelowi na jednoczesną analizę cech o różnej skali czasowej (np. szybkich załamków QRS oraz wolniejszych fal T).

2.3 Transformacja danych i kanały pochodne

Z uwagi na fakt, że surowe sygnały EKG są podatne na różnego rodzaju zakłócenia (szum mięśniowy, pływanie linii izoelektrycznej), opracowano wieloetapowy proces oczyszczania i transformacji danych:

1. **Filtracja pasmowo-przepustowa (Bandpass Filter):** Zastosowano filtr Butterwortha 4. rzędu o pasmie przenoszenia **0.5 – 40 Hz**. Pozwala to na usunięcie zakłóceń niskoczęstotliwościowych (oddechowych) oraz wysokoczęstotliwościowych szumów sieciowych i mięśniowych, zachowując jednocześnie kluczowe pasmo diagnostyczne zespołu QRS.
2. **Prycinanie amplitudy (Amplitude Clipping):** W celu eliminacji sygnałów o nienaturalnie wysokich wartościach (np. wynikających z chwilowego odpięcia elektrody), zaimplementowano *Clipper*. Wartości sygnału przekraczające **5 odchyleń standardowych** (5σ) są przycinane, co stabilizuje proces uczenia sieci neuronowych.
3. **Standaryzacja globalna:** Sygnał został poddany standaryzacji (*Z-score normalization*) na poziomie każdego odprowadzenia z osobna. Parametry (μ, σ) zostały wyznaczone wyłącznie na zbiorze treningowym i zaaplikowane do pozostałych zbiorów, co zapewnia spójną skalę danych wejściowych.
4. **Generowanie kanałów pochodnych:** Każdy 12-kanałowy rekord rozszerzono o pierwszą pochodną czasową (różnice sąsiednich próbek). Zastosowanie pochodnej pozwala modelowi lepiej reagować na dynamikę zmian sygnału, co jest kluczowe w detekcji ostrych załamków R. Ostateczny tensor wejściowy ma wymiar **24 kanały × 1000 próbek**.

2.4 Strategia Augmentacji Danych

Aby zwiększyć zdolności generalizacyjne modeli i ograniczyć zjawisko przeuczenia, w fazie treningu zastosowano dynamiczną augmentację sygnału (klasa `ECGAugmenter`). Wykorzystano następujące techniki:

1. **Gaussian Noise:** Dodawanie losowego szumu o rozkładzie normalnym ($\sigma = 0.01$). Pozwala to modelowi uodpornić się na zakłócenia elektryczne aparatury pomiarowej.
2. **Random Scaling:** Losowe skalowanie amplitudy sygnału w zakresie [0.9, 1.1]. Technika ta symuluje różnice w przewodnictwie tkanek pacjentów oraz różną czułość elektrod.

3. **Time Shifting:** Losowe przesunięcie sygnału w czasie o maksymalnie ± 50 próbek (0.5 sekundy). Zmusza to sieć do nauki cech niezależnych od dokładnego momentu wystąpienia zespołu QRS w oknie czasowym (niezmienność względem przesunięcia).

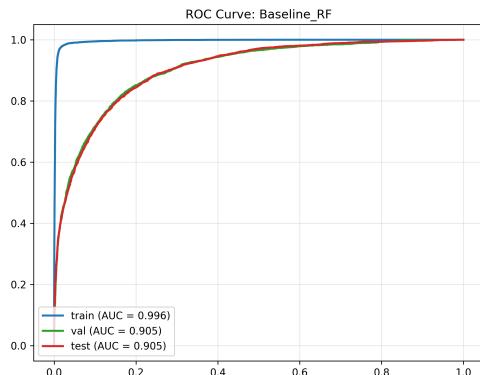
Warto podkreślić, że augmentacja była stosowana **wyłącznie na zbiorze treningowym**. Zbiory walidacyjny i testowy pozostały nienaruszone, aby zapewnić rzetelną ocenę realnej wydajności modeli na czystych danych klinicznych.

3 Eksperymenty i analiza wyników

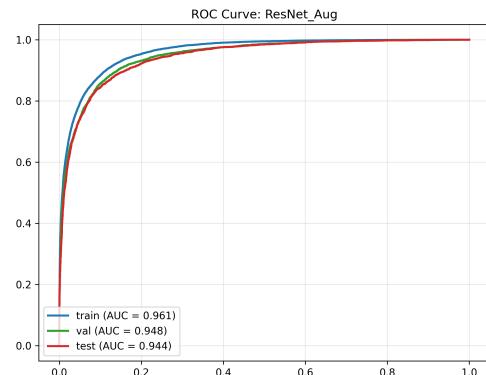
W ramach prac przeprowadzono trzy główne etapy eksperymentalne, mające na celu weryfikację postawionych hipotez badawczych dotyczących skuteczności metod głębokiego uczenia w klasyfikacji sygnałów EKG.

3.1 Eksperiment 1: Baseline vs Deep Learning

Pierwszy eksperyment dotyczył porównania klasycznego podejścia opartego na inżynierii cech (Random Forest) z modelem operującym na surowym sygnale (ResNet1D Augmented). Jak pokazują krzywe ROC, modele głębokie wykazują znacznie wyższą stabilność i czułość w całym zakresie progu decyzyjnego.



Rysunek 5: Krzywa ROC - Model Baseline (RF).



Rysunek 6: Krzywa ROC - ResNet1D Augmented.

W celu potwierdzenia istotności tych różnic wykonano **test statystyczny McNeaura**. Uzyskana wartość $p = 1,42 \times 10^{-14}$ pozwala jednoznacznie stwierdzić, że przewaga sieci neuronowych nad lasem losowym nie jest dziełem przypadku, lecz wynika z lepszej zdolności do ekstrakcji nieliniowych cech morfologicznych z zapisu EKG.

3.2 Eksperiment 2: Porównanie architektur ResNet i Inception-Time

Eksperyment polegał na zestawieniu dwóch wiodących architektur splotowych. Wyniki zestawione w tabeli 2 wskazują, że obie architektury osiągają niemal identyczny poziom *ROC-AUC* (0,94). Wybór między nimi jest trudny, gdyż InceptionTime lepiej radzi sobie

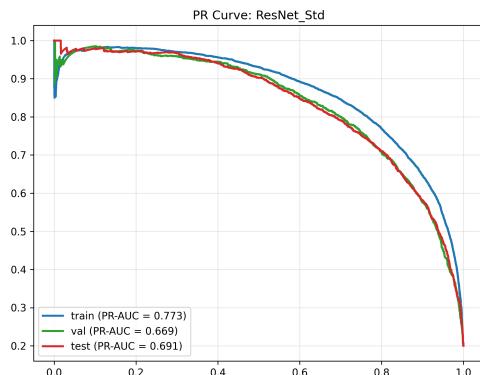
z klasą STTC, natomiast ResNet wykazuje lepszą precyzję w detekcji zmian zawałowych (MI).

Tabela 2: Zbiorcze zestawienie metryk dla wszystkich modeli.

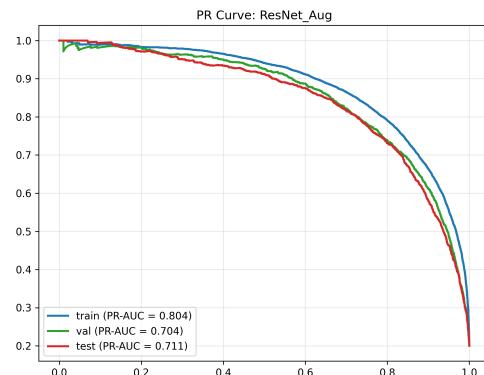
Model	Accuracy	F1-Macro	ROC-AUC	PR-AUC
Baseline (Random Forest)	0,667	0,499	0,904	0,577
ResNet1D (Standard)	0,752	0,644	0,942	0,691
ResNet1D (Augmented)	0,761	0,635	0,944	0,711
InceptionTime (Standard)	0,751	0,631	0,946	0,702
InceptionTime (Augmented)	0,758	0,643	0,943	0,698

3.3 Eksperyment 3: Wpływ augmentacji i pochodnych sygnału

Ostatni etap badań skupił się na weryfikacji wpływu zaawansowanego przygotowania danych (dodanie 12 kanałów pochodnych) oraz dynamicznej augmentacji na jakość klasyfikacji. Szczegółową różnicę w zachowaniu modeli najlepiej oddają krzywe *Precision-Recall* (Rysunek ??).



Rysunek 7: Krzywa PR - ResNet Standard (PR-AUC: 0,691).



Rysunek 8: Krzywa PR - ResNet Augmented (PR-AUC: 0,711).

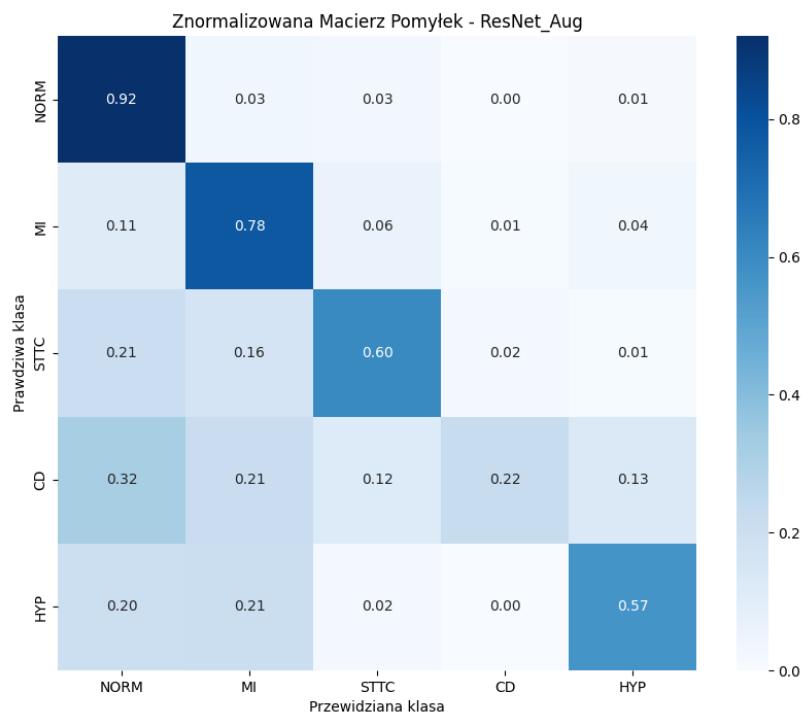
Analiza wyników prowadzi do interesujących wniosków dotyczących wpływu augmentacji na poszczególne architektury:

- **InceptionTime:** W przypadku tej architektury różnica w metrykach *PR-AUC* między wersją standardową (0,702) a augmentowaną (0,698) jest minimalna. Tak niewielkie odchylenie nie pozwala na negację tezy o przydatności augmentacji, sugerując raczej, że model ten posiada silną wewnętrzną zdolność do generalizacji, a dodatkowy szum i przesunięcia nie zaburzają wyuczonych wzorców.
- **ResNet1D:** Tutaj wyniki przemawiają zdecydowanie na korzyść zaproponowanego potoku przetwarzania. Wzrost wskaźnika *PR-AUC* o **2 punkty procentowe** (z 0,691 na 0,711) stanowi silny argument potwierdzający, że połączenie pochodnych sygnału z augmentacją czasowo-amplitudową pozwala sieci ResNet na skuteczniejsze wyodrębnienie kluczowych cech diagnostycznych.

Ostatecznie, ze względu na najwyższą precyzję ($PR-AUC$) oraz najbardziej klarowny zysk z zastosowanych transformacji, to model **ResNet1D Augmented** został wybrany jako finalne rozwiązanie, które poddano szczegółowej analizie interpretowalności.

3.4 Analiza jakości predykcji na podstawie macierzy pomyłek

Do szczegółowej oceny zdolności dyskryminacyjnych modelu wykorzystano znormalizowaną macierz pomyłek (Rysunek 9), która ujawnia stopień separacji poszczególnych jednostek chorobowych.



Rysunek 9: Znormalizowana macierz pomyłek dla najlepszego modelu (ResNet1D Augmented).

Analiza wyników pozwala na sformułowanie następujących wniosków:

- **Wysoka precyzja klasy NORM:** Model poprawnie identyfikuje 92% pacjentów zdrowych. Bardzo niski odsetek pomyłek w tej klasie świadczy o wysokiej specyficzności modelu, co jest kluczowe w badaniach przesiewowych.
- **Problemy z klasyfikacją CD:** Największym wyzwaniem dla modelu okazała się klasa zaburzeń przewodzenia (CD), gdzie poprawność wynosi jedynie 22%. Aż 32% tych przypadków jest błędnie klasyfikowanych jako **NORM**, a 21% jako **MI**. Sugeruje to, że lokalne zmiany przewodzenia są zbyt subtelne dla modelu lub ich wzorce nakładają się na inne patologie w obrębie 10-sekundowego zapisu.
- **Relacja MI i HYP:** Klasy zawałowa (MI - 78%) oraz przerostowa (HYP - 57%) są rozpoznawane ze stosunkowo dobrą skutecznością. Warto zauważyć, że błędy w

klasie **HYP** rozkładają się niemal równomiernie między **NORM** a **MI**, co może wynikać z wysokich amplitud załamków, które występują zarówno w ostrych stanach niedokrwieniennych, jak i w przerostach.

- **Niejasność klasy STTC:** Model poprawnie klasyfikuje 60% przypadków **STTC**, jednak często myli je z **NORM** (21%) i **MI** (16%), co znajduje uzasadnienie w medycznej specyfice tych zmian – są one często niespecyficzne i mogą zapowiadać zawał lub być wariantem normy.

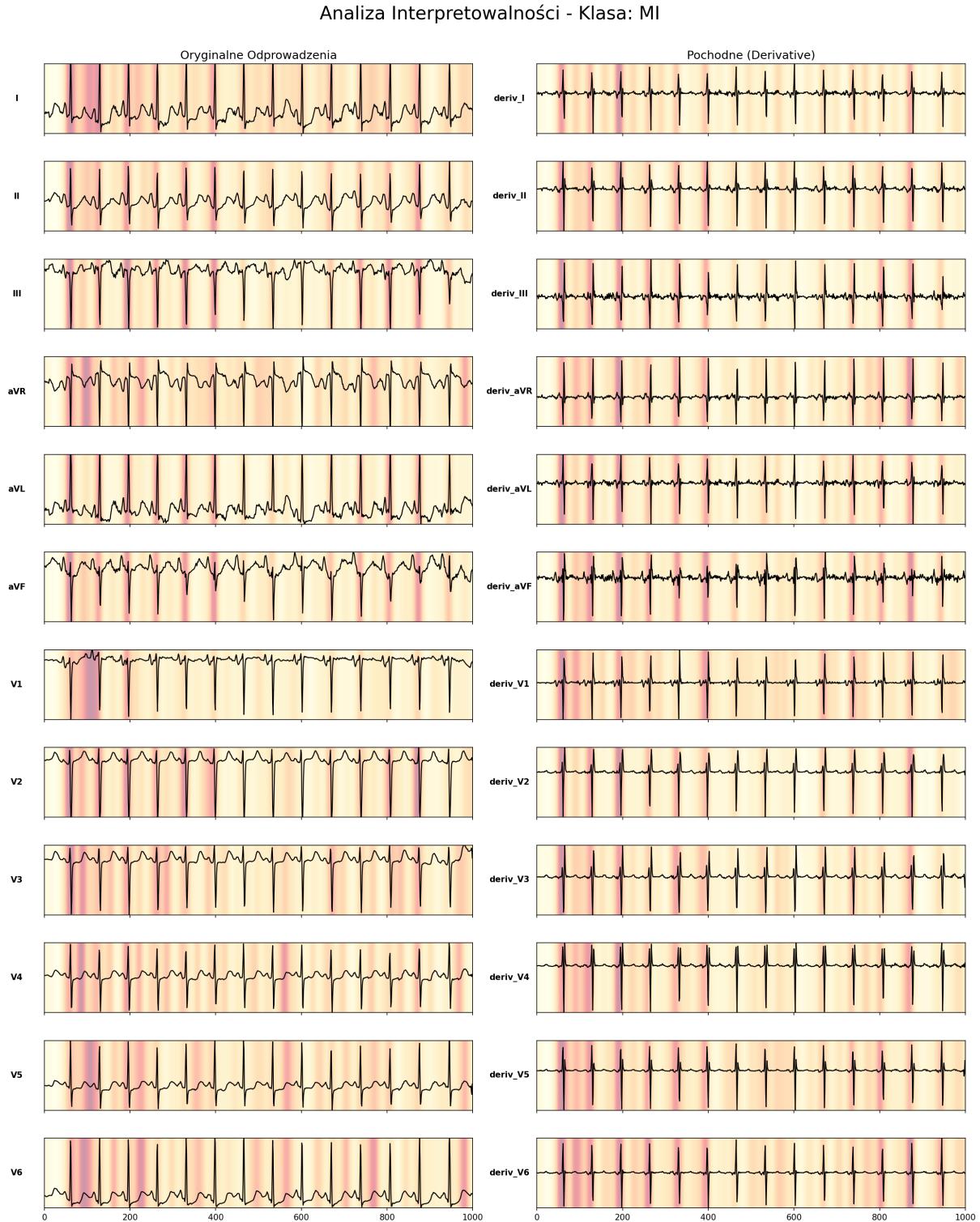
4 Interpretowalność modelu – Saliency Maps

W diagnostyce medycznej wspieranej przez sztuczną inteligencję kluczowym aspektem jest nie tylko skuteczność metryczna, ale również możliwość interpretacji decyzji modelu. W niniejszym projekcie zastosowano metodę **Saliency Maps**, która bazuje na obliczaniu gradientów funkcji wyjściowej względem wejścia. Pozwala to na wizualizację „obszarów zainteresowania” sieci neuronowej i weryfikację, czy model podejmuje decyzje w oparciu o klinicznie istotne cechy morfologiczne sygnału EKG, czy też opiera się na zakłóceniami (tzw. *Clever Hans effect*).

Do analizy wybrano dwa reprezentatywne przypadki dla najlepszego modelu (*ResNet1D Augmented*), wykorzystującego rozszerzoną reprezentację 24-kanałową.

4.1 Przypadek 1: Prawidłowa detekcja zawału (Klasa MI)

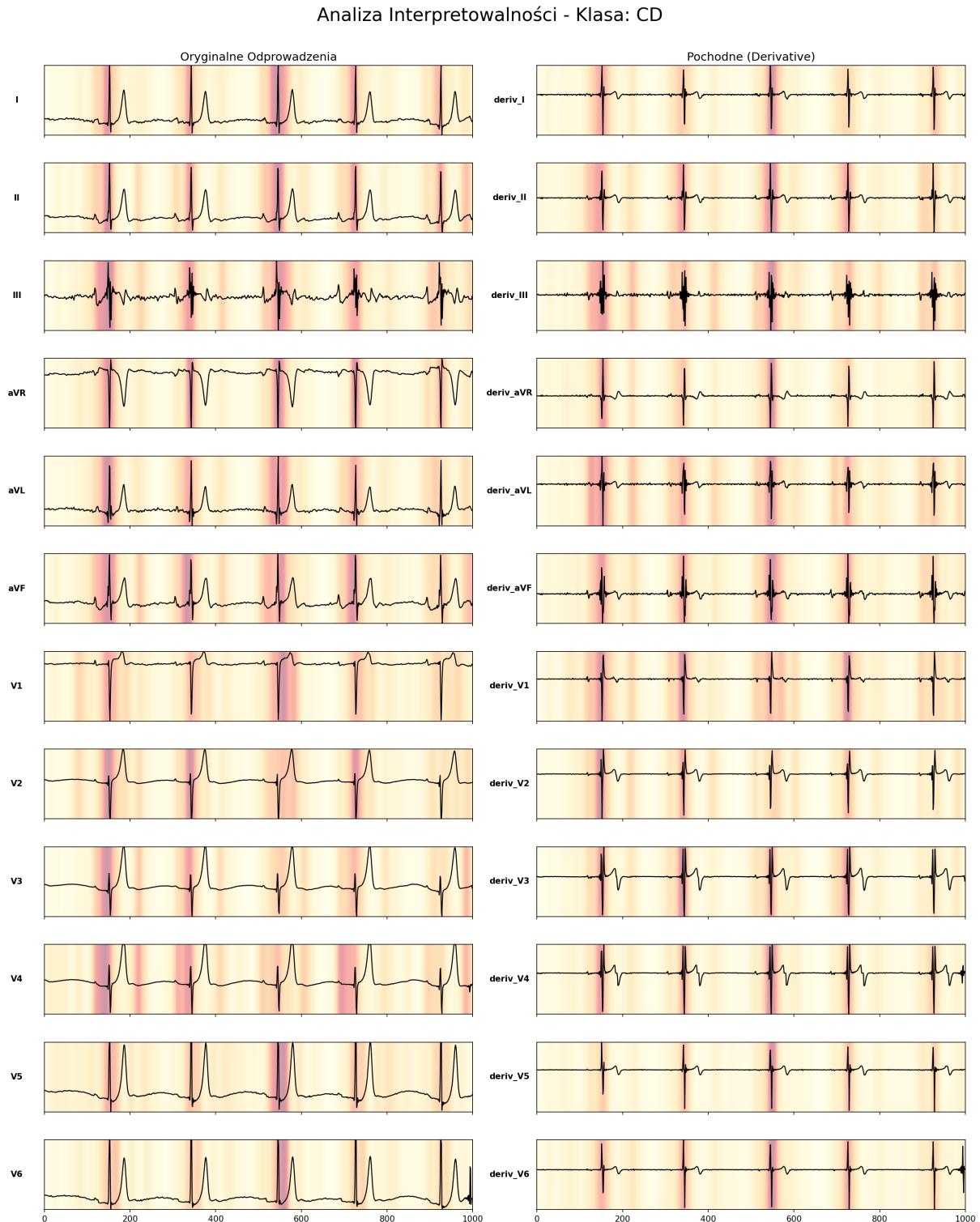
W tym eksperymencie poddano analizie rekord, który model poprawnie sklasyfikował jako zawał mięśnia sercowego (*Myocardial Infarction*). Klasa ta charakteryzuje się wysoką skutecznością w macierzy pomyłek (78%).



Rysunek 10: Mapa istotności dla klasy MI (24 kanały). Górnne 12 wykresów to surowe odprowadzenia, dolne 12 to ich pochodne czasowe.

4.2 Przypadek 2: Błędna klasyfikacja zaburzeń przewodzenia (Klasa CD)

Drugi przykład dotyczy rekordu z klasy zaburzeń przewodzenia (*Conduction Disturbance*), z którą model radził sobie najgorzej (22%).



Rysunek 11: Mapa istotności dla błędnie sklasyfikowanego przypadku CD (24 kanały). Widoczna rozproszona uwaga modelu.

5 Podsumowanie i wnioski

Niniejszy projekt miał na celu opracowanie i ewaluację zaawansowanych modeli głębokiego uczenia do automatycznej klasyfikacji sygnałów EKG ze zbioru PTB-XL. Przeprowadzone badania pozwoliły na sformułowanie następujących wniosków końcowych:

1. **Wyższość modeli głębokich nad klasycznymi:** Eksperymenty jednoznacznie wykazały, że sieci splotowe (ResNet, InceptionTime) przewyższają model bazowy *Random Forest*. Potwierdza to hipotezę, że automatyczna ekstrakcja cech z surowego sygnału jest skuteczniejsza niż ręczne definiowanie statystyk w przypadku tak złożonych danych medycznych.
2. **Wpływ augmentacji i kanałów pochodnych:** Zastosowanie autorskiego potoku przetwarzania (pochodne czasowe + augmentacja) przyniosło największe korzyści dla architektury **ResNet1D**, poprawiając wskaźnik *PR-AUC* o 2 punkty procentowe. Okazało się to kluczowe dla stabilności predykcji klas mniejszościowych.
3. **Specyfika klas diagnostycznych:** Model wykazał doskonałą zdolność rozpoznawania wyników prawidłowych (*NORM* - 92%), co predysponuje go do roli narzędzia przesiewowego. Największym wyzwaniem pozostają zaburzenia przewodzenia (*CD*), które wymagają prawdopodobnie dłuższych okien czasowych lub większej liczby danych treningowych do poprawnej identyfikacji.
4. **Znaczenie interpretowalności:** Wykorzystanie map istotności (*Saliency Maps*) pozwoliło na „otwarcie czarnej skrzynki” modelu. Potwierdzono, że w przypadkach zawałowych (*MI*) sieć skupia uwagę na klinicznie istotnych segmentach ST, co buduje zaufanie do decyzji podejmowanych przez algorytm.