



**UNIWERSYTET
MIKOŁAJA KOPERNIKA
W TORUNIU**

Collegium Medicum
im. Ludwika Rydygiera w Bydgoszczy

Bydgoszcz 2026 r.



mgr farm. Łukasz Ledziński

Implementacja metod sztucznej inteligencji dla ułatwienia podejmowania decyzji klinicznych w kardiologii

**Rozprawa na stopień doktora nauk medycznych i nauk o zdrowiu
*w dyscyplinie nauki farmaceutyczne***

Promotor: prof. dr hab. n. med. Grzegorz Grzešek

Katedra Kardiologii i Farmakologii Klinicznej

Kierownik Katedry: prof. dr hab. n. med. Grzegorz Grzešek

Bydgoszcz 2026 r.

Najbliższym

Dziękuję wszystkim, którzy byli dla mnie wsparciem, inspiracją i motywacją do realizacji tej pracy. Szczególnie za ich cierpliwość, zaangażowanie oraz okazaną życzliwość.

Dziękuję Wam z całego serca.

SPIS TREŚCI

SPIS TREŚCI.....	1
WYKAZ SKRÓTÓW.....	3
WYKAZ PUBLIKACJI WCHODZĄCYCH W SKŁAD PRACY.....	5
WYKAZ RYCIN.....	7
WYKAZ RÓWNAŃ.....	8
WYKAZ TABEL.....	9
1. WSTĘP.....	10
2. ZAŁOŻENIA I CELE PRACY	16
2.1. CEL GŁÓWNY PRACY:.....	16
2.2. CELE SZCZEGÓŁOWE:.....	16
3. MATERIAŁY I METODY	17
3.1. WSPÓLNE DLA WSZYSTKICH PUBLIKACJI	17
3.1.1. Algorytmy oparte o drzewa decyzyjne	17
3.1.2. Metryki.....	19
3.1.3. Wartości Shapley'a.....	21
3.1.4. LASSO	22
3.1.5. Analiza statystyczna.....	23
3.1.6. Optymalizacja hiperparametrów	24
3.1.7. Środowisko obliczeniowe.....	26
3.2. PUBLIKACJA I.....	26
3.3. PUBLIKACJA II	27
3.3.1. Źródło danych.....	27
3.3.2. Kryteria włączenia do badania.....	27
3.3.3. Przygotowanie danych.....	28
3.3.4. Braki danych.....	29
3.3.5. Budowa modeli uczenia maszynowego.....	29
3.4. PUBLIKACJA III.....	30
3.4.1. Źródło danych.....	30
3.4.2. Kryteria włączenia do badania.....	30
3.4.3. Przygotowanie danych.....	31
3.4.4. Braki w danych.....	31
3.4.5. Budowa modeli uczenia maszynowego	32
4. WYNIKI.....	33
4.1. PUBLIKACJA I.....	33
4.2. PUBLIKACJA II	38

4.2.1.	<i>Analiza statystyczna</i>	38
4.2.2.	<i>Metryki</i>	42
4.2.3.	<i>SHAP</i>	44
4.3.	PUBLIKACJA III	46
4.3.1.	<i>Analiza statystyczna</i>	46
4.3.2.	<i>Metryki</i>	51
4.3.3.	<i>SHAP</i>	52
5.	DYSKUSJA	55
5.1.	PODSUMOWANIE	69
5.2.	OGRANICZENIA	69
5.3.	OGRANICZENIA METODOLOGICZNE	70
5.4.	SZANSE APLIKACYJNE	75
6.	WNIOSKI	76
7.	PIŚMIENNICTWO	77
8.	STRESZCZENIE	90
9.	ABSTRACT	92
10.	ZAŁĄCZNIKI	94
10.1.	PUBLIKACJA I.....	94
10.2.	PUBLIKACJA II	116
10.3.	PUBLIKACJA III	131

WYKAZ SKRÓTÓW

AdaBoost – Adaptive Boosting

ADASYN – ang. *Adaptive Synthetic Sampling*, metoda nadpróbkowywania

AF – migotanie przedsionków

AI – sztuczna inteligencja

ANN – sztuczna sieć neuronowa

BNP-PL – polski rejestr pacjentów z rozpoznaniem nadciśnieniem płucnym

CNN – konwolucyjna sieć neuronowa

CTEPH – d przewlekłe zakrzepowo-zatorowe nadciśnienie płucne

DOAC – bezpośrednie doustne leki przeciwkrzepliwe

ESC – Europejskie Towarzystwo Kardiologiczne

FN – fałszywie negatywne

FP – fałszywie pozytywne

FPR – ang. *False Positive Rate* – 1 – specyficzność, stosunek liczby błędnie zaklasyfikowanych negatywnych przypadków do wszystkich rzeczywistych negatywnych przykładów.

Grad-CAM – ang. *Gradient-weighted class activation mapping*, metoda należąca do XAI

LASSO – ang. *Least Absolute Shrinkage and Selection Operator*

LightGBM – Light Gradient Boosting Machine

LLM – duży model językowy

LR – regresja logistyczna

MCC – współczynnik korelacji Matthews

MICE – ang. *Multiple Imputation by Chained Equations*

MIMIC IV – baza danych *Medical Information Mart for Intensive Care*

ML – uczenie maszynowe

NLP – przetwarzanie języka naturalnego
PAH – płucne nadciśnienie tętnicze
RF – las losowy
RNN – rekurencyjna sieć neuronowa
SFE – sekwencyjna eliminacja cech
SHAP – wartości Shapley’a
SMOTE – ang. *Synthetic Minority Over-sampling Technique*, metoda nadpróbkowania
SMOTETomek – połączenie techniki SMOTE i podpróbkowania w oparciu o łączy Tomka
sPAP – skurczowe ciśnienie w tętnicy płucnej
SVM – maszyna wektorów nośnych
TAPSE – amplituda skurczowego przemieszczania pierścienia zastawki trójdzielnej
TN – prawdziwie negatywne
TP – prawdziwie pozytywne
TPR – ang. *True Positive Rate* –inaczej czułość lub recall
XAI – wytłumaczalna sztuczna inteligencja
XGBoost – eXtreme Gradient Boosting

WYKAZ PUBLIKACJI WCHODZĄCYCH W SKŁAD PRACY

Publikacja I:

(praca pogładowa)

Łukasz Ledziński; Grzegorz Grzešek

Artificial Intelligence Technologies in Cardiology

J. Cardiovasc. Dev. Dis. 2023, 10, 202.

<https://doi.org/10.3390/jcdd10050202>

IF: 2,3

Punktacja MNiSW: 20

Publikacja II:

(praca oryginalna)

Łukasz Ledziński; Elżbieta Grzešek; Małgorzata Ledzińska; Grzegorz Grzešek

Predicting Mortality in Atrial Fibrillation Patients Treated with Direct Oral Anticoagulants: A Machine Learning Study Based on the MIMIC-IV Database

J. Clin. Med. 2025, 14, 3697.

<https://doi.org/10.3390/jcm14113697>

IF: 2,9

Punktacja MNiSW: 140

Publikacja III:

(praca oryginalna)

Łukasz Ledziński; Grzegorz Grześk; Michał Ziołkowski; Marcin Waligóra; Marcin Kurzyna; Tatiana Mularek-Kubzdela; Anna Smukowska-Gorynia; Ilona Skoczyła; Łukasz Chrzanowski; Piotr Błaszczak; Miłosz Jaguszewski; Beata Kuśmierczyk-Droszcz; Katarzyna Ptaszyńska; Katarzyna Mizia-Stec; Ewa Malinowska; Małgorzata Peregud-Pogorzelska; Ewa Lewicka; Michał Tomaszewski; Wojciech Jacheć; Michał Floreczyk; Ewa Mroczek; Zbigniew Gąsior; Agnieszka Pawlak; Katarzyna Betkier-Lipińska; Piotr Pruszczyk; Olga Dzikowska-Diduch; Katarzyna Widejko; Judyta Winowska-Józwa; Grzegorz Kopec
Personalized Medicine in Pulmonary Arterial Hypertension: Utilizing Artificial Intelligence for Death Prevention

J. Clin. Med. 2025, 14, 8325.

<https://doi.org/10.3390/jcm 14238325>

IF: 2,9

Punktacja MNiSW: 140

WYKAZ RYCIN

Rycina 1. Wykres SHAP globalny	45
Rycina 2. Wykres SHAP lokalny dla pacjenta, u którego model prawidłowo przewidział zgon.	46
Rycina 3. Wykres SHAP globalny	53
Rycina 4. Wykres SHAP lokalny dla pacjenta, u którego model prawidłowo przewidział zgon.	54

WYKAZ RÓWNAŃ

Równanie 1. Dokładność	19
Równanie 2. Swoistość	19
Równanie 3. Czułość	20
Równanie 4. Współczynnik korelacji Matthews'a.....	20
Równanie 5. AUC.....	20
Równanie 6. Wartość Shapley'a.....	22
Równanie 7. LASSO.....	23
Równanie 8. Optymalizacja HyperOpt.....	25

WYKAZ TABEL

Tabela 1. Macierz błędów	21
Tabela 2. Przykłady zastosowań AI/ML w kardiologii	33
Tabela 3. Analiza statystyczna 27 cech użytych do budowy modelu.....	39
Tabela 4. Metryki wydajności na zbiorze treningowym.....	42
Tabela 5. Metryki wydajności na zbiorze testowym	43
Tabela 6. Analiza statystyczna 17 cech użytych do budowy modelu.....	49
Tabela 7. Metryki wydajności na zbiorze testowym	51
Tabela 8. Macierz błędów dla zbioru testowego	52
Tabela 9. Porównanie wyników uzyskanych przez model z Publikacji II z innymi dostępnymi w literaturze	60
Tabela 10. Porównanie wyników uzyskanych przez model z Publikacji III z innymi dostępnymi w literaturze.....	64
Tabela 11. Zestawienie parametrów modeli z publikacji II i III	67

1. Wstęp

Współczesna medycyna znajduje się w przełomowym momencie rozwoju, mierząc się z bezprecedensowym wyzwaniem, jakim jest efektywne zarządzanie i wykorzystanie ogromnych ilości danych generowanych każdego dnia w praktyce klinicznej. Zgodnie z prognozami International Data Corporation, w roku 2025 globalna ilość danych osiągnęła poziom 175 zettabajtów, a tempo ich przyrostu wykazuje charakter wykładniczy¹. Szacuje się, że przeciętny użytkownik wchodzi w cyfrową interakcję raz na 18 sekund². Źródła tych danych są niezwykle zróżnicowane – obejmują nie tylko tradycyjną dokumentację medyczną, wyniki badań laboratoryjnych i obrazowych, lecz także dane pochodzące z urządzeń noszonych przez pacjentów, aplikacji mobilnych wspierających terapię oraz systemów monitorujących stan zdrowia w czasie rzeczywistym. Tak dynamiczny wzrost informacji, określany akronimem „4V” (volume, velocity, veracity, variety – objętość, szybkość napływu, wiarygodność i różnorodność), sprawia, że klasyczne metody zarządzania, przetwarzania, analizy oraz zapewnienia bezpieczeństwa danych stają się coraz mniej wydolne i nieadekwatne do aktualnych potrzeb³.

W odpowiedzi na te wyzwania, coraz większe nadzieje pokłada się w zaawansowanych narzędziach analitycznych, takich jak sztuczna inteligencja (AI), uczenie maszynowe (ML) oraz w ostatnich latach duże modele językowe (LLM). Technologie te umożliwiają analizę złożonych, wielowymiarowych zbiorów danych, identyfikację ukrytych wzorców oraz wspieranie procesu podejmowania decyzji klinicznych w sposób spersonalizowany. Sztuczna inteligencja, rozumiana jako systemy naśladujące ludzką inteligencję, obejmuje zarówno algorytmy uczące się na podstawie danych historycznych, jak i systemy zdolne do adaptacyjnego

podejmowania decyzji w oparciu o nowe informacje. Najnowsze systemy oparte o strukturę agentową dają możliwość pełnej integracji w środowisko pracy i delegowania codziennych zadań. Zainteresowanie AI w medycynie sięga lat 50. XX wieku, kiedy Alan Turing sformułował fundamentalne pytanie o możliwość myślenia przez maszyny, a John McCarthy wprowadził termin „sztuczna inteligencja”^{4,5}. Współcześnie AI i ML znajdują szerokie zastosowanie w analizie obrazów medycznych, przetwarzaniu języka naturalnego, ocenie danych z urządzeń noszonych oraz w budowie zaawansowanych systemów wspomagania decyzji klinicznych, co stanowi jeden z najważniejszych trendów badawczych ostatnich lat.

Kardiologia jest dziedziną szczególnie predysponowaną do wdrażania nowoczesnych technologii analitycznych. Choroby sercowo-naczyniowe, takie jak choroba wieńcowa, niewydolność serca, arytmie czy wady zastawkowe, pozostają główną przyczyną zgonów na świecie stanowiąc blisko jedną trzecią wszystkich zgonów w roku 2023⁶. Diagnostyka i leczenie tych schorzeń wymagają integracji danych pochodzących z wielu źródeł oraz zaawansowanej wiedzy specjalistycznej. Dodatkowo, duża liczba pacjentów oraz złożoność przypadków klinicznych generują potrzebę opracowania skutecznych narzędzi diagnostycznych, przesiewowych i wspierających codzienną praktykę kliniczną. Współczesna kardiologia korzysta z zaawansowanych technik obrazowania, takich jak echokardiografia, rezonans magnetyczny serca czy tomografia komputerowa, które dostarczają szczegółowych informacji o budowie i funkcji serca. Coraz większe znaczenie mają także dane z urządzeń wszczepialnych, a rosnąca liczba pacjentów korzysta z

mobilnych urządzeń monitorujących rytm serca i inne parametry życiowe, stwarza to kolejne bogate źródło danych.

W ostatnich latach obserwuje się intensyfikację wykorzystania AI i ML w kardiologii, zarówno w zakresie poprawy dokładności i efektywności diagnostyki, jak i leczenia oraz monitorowania pacjentów. Algorytmy analizy obrazów medycznych umożliwiają szybkie i precyzyjne wykrywanie nieprawidłowości, natomiast modele predykcyjne pozwalają na identyfikację pacjentów wysokiego ryzyka oraz prognozowanie rozwoju powikłań^{7–14}. Systemy wspomagania decyzji klinicznych oparte na AI wspierają lekarzy w wyborze optymalnej terapii, a urządzenia noszone przez pacjentów umożliwiają wczesne wykrywanie zaburzeń rytmu serca, takich jak bradykardia czy migotanie przedsionków, monitorowanie spadków saturacji krwi oraz zmienności tętna. Rozwój technologii AI otwiera nowe perspektywy dla kardiologii, umożliwiając dalszą personalizację opieki oraz poprawę wyników leczenia.

Jednym z częstszych schorzeń kardiologicznych jest migotanie przedsionków (AF), będące najczęstszą arytmia nadkomorową, której częstość występowania wzrasta wraz z wiekiem populacji. Zgodnie z danymi Global Burden of Disease w roku 2023 współczynnik zapadalności i chorobowości, z powodu migotania i trzepotania przedsionków, na 100 tysięcy mieszkańców wynosił na świecie odpowiednio 62.26 (95% CI 49.39 – 78.67) i 732.04 (95% CI 576.37 – 902.18), a współczynnik zgonów 4.68 (95% CI 3.95 – 5.25)¹⁵. AF wiąże się ze znacznym wzrostem ryzyka udaru mózgu, niewydolności serca oraz zgonu, co istotnie wpływa na jakość i długość życia pacjentów. Kompleksowa opieka nad pacjentem z AF skupia się na kilku filarach: zarządzaniu chorobami współistniejącymi i czynnikami ryzyka, prewencji udarów mózgu i powikłań zakrzepowo-

zatorowych, redukcji objawów poprzez kontrolę częstości i rytmu serca, ewaluacji i dynamicznej ponownej ocenie. Kluczową rolę w leczeniu AF odgrywają doustne antykoagulanty niebędące antagonistami witaminy K (DOAC), takie jak dabigatran, rywaroksaban, apiksaban i edoksaban, umożliwiające prewencję incydentów zakrzepowo-zatorowych ¹⁶. W porównaniu do tradycyjnych antagonistów witaminy K, DOAC cechują się przewidywalną farmakokinetyką, mniejszą liczbą interakcji oraz niższym ryzykiem poważnych krwawień, co czyni je preferowaną opcją terapeutyczną w większości przypadków AF. Liczne badania kliniczne i obserwacyjne potwierdziły skuteczność i bezpieczeństwo DOAC w prewencji udarów oraz powikłań zakrzepowo-zatorowych.

Pomimo postępu w leczeniu, przewidywanie ryzyka powikłań i śmiertelności u pacjentów z AF pozostaje wyzwaniem. Aktualnie stosowane skale ryzyka, takie jak CHA₂DS₂-VASc, HATCH czy HAVOC, wykazują ograniczoną wartość prognostyczną, zwłaszcza w zakresie przewidywania śmiertelności ^{17–20}. W związku z tym coraz częściej podejmowane są próby wykorzystania narzędzi uczenia maszynowego do poprawy predykcji ryzyka w tej populacji. Algorytmy ML umożliwiają analizę złożonych zależności pomiędzy wieloma czynnikami ryzyka, takimi jak wiek, choroby współistniejące, wyniki badań laboratoryjnych i obrazowych, a także dane z wywiadu i stylu życia pacjenta, co pozwala na bardziej precyzyjne dostosowanie terapii i poprawę rokowania ^{21–23}.

Tętnicze nadciśnienie płucne (PAH) stanowi rzadką, wieloczynnikową chorobę sercowo-naczyniową, charakteryzującą się wzrostem oporu w tętnicach płucnych, prowadzącym do przeciążenia i niewydolności prawej komory serca oraz znacznie zwiększonego ryzyka

zgonu. Wczesna diagnostyka i wdrożenie odpowiedniego leczenia są kluczowe dla poprawy rokowania. Terapia PAH obejmuje m.in. stosowanie blokerów kanału wapnia, antagonistów receptora endotelinowego, inhibitorów fosfodiesterazy typu 5, analogów prostacyklin, agonistów receptora prostacyklin i stymulatora rozpuszczalnej cyklicznej guanylowej oraz terapie wspomagające, leczenie interwencyjne i transplantacje ^{24,25}. Zgodnie z danymi Global Burden of Disease w roku 2023 współczynnik zapadalności i chorobowości, z powodu tętniczego nadciśnienia płucnego, na 100 tysięcy mieszkańców wynosił na świecie odpowiednio 0.56 (95% CI 0.45 – 0.68) i 2.45 (95% CI 1.98 – 3.04), a współczynnik zgonów 0.28 (95% CI 0.22 – 0.37) ¹⁵. Skale ryzyka, takie jak REVEAL ^{26,27}, COMPERA ²⁸ i oparte o wytyczne ESC ^{29,30}, są wykorzystywane do oceny rokowania oraz stratyfikacji pacjentów do grup ryzyka, jednak ich skuteczność jest ograniczona przez złożoność choroby i różnorodność czynników ryzyka ^{24,31}. W ostatnich latach coraz większą nadzieję pokłada się w modelach uczenia maszynowego, które pozwalają na integrację danych klinicznych, laboratoryjnych, obrazowych i biochemicznych, umożliwiając bardziej precyzyjną identyfikację pacjentów wysokiego ryzyka i personalizację terapii ³². Badania wykazały, że podejścia oparte na AI mogą przewyższać tradycyjne metody statystyczne w predykcji niekorzystnych zdarzeń i śmiertelności w PAH.

Wdrożenie AI w kardiologii wiąże się jednak z licznymi i złożonymi wyzwaniami. Kluczowe znaczenie ma zapewnienie wysokiej jakości, kompletności i bezpieczeństwa danych, które stanowią podstawę działania algorytmów ML. Problemy związane z brakującymi danymi, błędami w dokumentacji czy niejednorodnością źródeł mogą prowadzić do błędnych predykcji i ograniczać użyteczność modeli. Dodatkowo, istotne są kwestie

etyczne i prawne, obejmujące ochronę prywatności pacjentów, odpowiedzialność za decyzje podejmowane na podstawie rekomendacji AI oraz konieczność spełnienia wymogów regulacyjnych^{33–38}. Niezwykle ważne jest także zapewnienie transparentności i interpretowalności modeli, aby lekarze mogli zrozumieć i zaufać rekomendacjom generowanym przez AI. Skuteczna integracja AI z codzienną praktyką kliniczną wymaga ścisłej współpracy interdyscyplinarnej, obejmującej zarówno ekspertów medycznych, jak i specjalistów z zakresu sztucznej inteligencji, analizy danych.

Podsumowując, dynamiczny rozwój sztucznej inteligencji i uczenia maszynowego otwiera nowe możliwości w kardiologii, zarówno w zakresie diagnostyki, jak i prognozowania oraz leczenia chorób sercowo-naczyniowych. Integracja zaawansowanych technik analitycznych z wiedzą kliniczną umożliwia tworzenie narzędzi wspierających personalizację opieki, wczesną identyfikację pacjentów wysokiego ryzyka oraz optymalizację decyzji terapeutycznych. Celem niniejszej rozprawy jest wykorzystanie nowoczesnych metod uczenia maszynowego do identyfikacji czynników ryzyka oraz budowy modelu predykcyjnego śmiertelności u pacjentów z PAH i AF, co może przyczynić się do poprawy jakości opieki i wyników leczenia w tych szczególnie wymagających populacjach.

2. Założenia i Cele pracy

2.1. Cel główny pracy:

Celem pracy było opracowanie modelu predykcyjnego z wykorzystaniem algorytmów sztucznej inteligencji, wspierającego codzienną praktykę kliniczną w opiece nad pacjentami z chorobami sercowo-naczyniowymi.

2.2. Cele szczegółowe:

Celem szczegółowym pracy było zbadanie możliwości zastosowania algorytmów uczenia maszynowego do przewidywania ryzyka zgonu pacjentów oraz identyfikacji kluczowych czynników wpływających na predykcje modeli:

- 1) W grupie pacjentów chorujących na migotanie przedsionków leczonych doustnymi lekami przeciwkrzepliwymi
- 2) W grupie pacjentów chorujących na płucne nadciśnienie tętnicze w populacji krajowej

3. Materiały i Metody

3.1. Wspólne dla wszystkich publikacji

3.1.1. Algorytmy oparte o drzewa decyzyjne

W ramach przeprowadzonych analiz zastosowano algorytmy oparte na drzewach decyzyjnych, które są jednymi z najczęściej wykorzystywanych metod uczenia maszynowego w zadaniach klasyfikacyjnych i regresyjnych. Ich działanie polega na hierarchicznym podziale danych na podstawie wartości cech, co prowadzi do budowy struktury drzewa, w której liście reprezentują przewidywane klasy lub wartości. Chociaż pojedyncze drzewa decyzyjne są łatwe w interpretacji, ich główną wadą jest podatność na przeuczenie, szczególnie w przypadku złożonych zbiorów danych. Aby temu zapobiec, stosuje się metody zespołowe, takie jak bagging i boosting, które łączą wiele drzew decyzyjnych w celu poprawy skuteczności i stabilności predykcji.

Bagging (*Bootstrap Aggregating*) to technika zespołowa, która polega na trenowaniu wielu niezależnych modeli na różnych podzbiorach danych, utworzonych za pomocą próbkowania z powtórzeniami (bootstrap). Wyniki poszczególnych modeli są następnie agregowane, na przykład poprzez głosowanie większościowe (dla klasyfikacji) lub uśrednianie (dla regresji). Bagging zmniejsza wariancję modelu, co czyni go bardziej odpornym na przeuczenie. Przykładem algorytmu wykorzystującego tę technikę jest las losowy (RF), który buduje wiele niezależnych drzew decyzyjnych, losując zarówno próbki danych, jak i podzbiory cech. RF cechuje się wysoką odpornością na przeuczenie oraz dobrą wydajnością w przypadku danych o dużej liczbie cech, jednak jego wadą jest stosunkowo wysoka złożoność obliczeniowa.

Boosting to technika zespołowa, która iteracyjnie buduje kolejne modele, przy czym każdy kolejny model stara się poprawić błędy popełnione przez poprzednie. W przeciwieństwie do baggingu, w którym modele są trenowane niezależnie, boosting tworzy modele w sposób sekwencyjny, co pozwala na stopniowe zmniejszanie błędów. Wzmocnienie gradientowe (*Gradient Boosting*) to szczególny rodzaj boostingu, w którym każdy kolejny model minimalizuje funkcję straty za pomocą metody gradientowej. Przykładami algorytmów opartych na wzmocnieniu gradientowym są XGBoost (Extreme Gradient Boosting) i LightGBM (Light Gradient Boosting Machine). XGBoost charakteryzuje się wysoką wydajnością dzięki optymalizacjom, takim jak równoległe przetwarzanie i regularyzacja, natomiast LightGBM został zoptymalizowany pod kątem szybkości i efektywności w przypadku dużych zbiorów danych, wykorzystując technikę histogramową i strategię "leaf-wise". Oba algorytmy są bardzo skuteczne, ale wymagają starannego strojenia hiperparametrów.

AdaBoost (Adaptive Boosting) to kolejny przykład algorytmu boostingu, w którym kolejne modele przypisują większe wagi próbom błędnie sklasyfikowanym przez poprzednie. Wynik końcowy uzyskiwany jest na podstawie ważonego głosowania. AdaBoost jest prosty w implementacji i dobrze radzi sobie z danymi o niskim poziomie szumu, jednak jego skuteczność może być ograniczona w przypadku bardzo złożonych zbiorów danych.

Zastosowanie algorytmów opartych na drzewach decyzyjnych, wykorzystujących techniki baggingu i boostingu, pozwoliło na uzyskanie wysokiej skuteczności predykcji oraz elastyczności w analizie danych. Ich zdolność do pracy z różnorodnymi typami danych oraz obsługi

brakujących wartości czyni je szczególnie przydatnymi w analizach medycznych. W pracy szczególny nacisk położono na interpretowalność wyników oraz optymalizację hiperparametrów, co pozwoliło na dostosowanie modeli do specyfiki analizowanych zbiorów danych.

3.1.2. Metryki

W celu oceny skuteczności opracowanych modeli predykcyjnych zastosowano zestaw standardowych metryk klasyfikacyjnych, które umożliwiły kompleksową analizę ich działania.

Dokładność (*Accuracy*) określa odsetek poprawnie sklasyfikowanych próbek w stosunku do wszystkich próbek, co pozwala na ocenę ogólnej jakości modelu. Jest to metryka intuicyjna w interpretacji, jednak w przypadku niezbalansowanych zbiorów danych może być myląca, ponieważ nie odzwierciedla wydajności modelu w odniesieniu do klasy mniejszościowej.

Równanie 1. Dokładność

$$\text{Dokładność} = \frac{TP + TN}{TP + TN + FP + FN}$$

Swoistość (*Specificity*) mierzy zdolność modelu do poprawnej klasyfikacji klasy negatywnej i jest szczególnie istotna w sytuacjach, w których kluczowe jest minimalizowanie liczby fałszywie pozytywnych wyników (błędy typu I).

Równanie 2. Swoistość

$$\text{Swoistość} = \frac{TN}{TN + FP}$$

Czułość (*Sensitivity*), znana również jako *recall*, ocenia zdolność modelu do poprawnej klasyfikacji klasy pozytywnej, co ma szczególne znaczenie w przypadkach, gdzie istotne jest ograniczenie liczby fałszywie negatywnych wyników (błędy typu II).

Równanie 3. Czulość

$$Czulość = \frac{TP}{TP + FN}$$

W pracy wykorzystano również współczynnik korelacji Matthews (MCC), który jest zrównoważoną miarą jakości klasyfikacji binarnej, uwzględniającą wszystkie elementy macierzy pomyłek. MCC jest szczególnie przydatny w analizie niezbalansowanych zbiorów danych, ponieważ dostarcza bardziej wiarygodnej oceny niż dokładność.

Równanie 4. Współczynnik korelacji Matthews

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

Dodatkowo, zastosowano pole pod krzywą ROC (ROC-AUC), które pozwala na ocenę zdolności modelu do rozróżniania między klasami. Wysoka wartość AUC wskazuje na dobrą zdolność modelu do przewidywania klasy pozytywnej, niezależnie od progu decyzyjnego. Zastosowanie tych metryk umożliwiło wszechstronną ocenę skuteczności modeli, uwzględniając zarówno ich ogólną dokładność, jak i wydajność w odniesieniu do poszczególnych klas.

Równanie 5. AUC

$$AUC = \int_0^1 TPR(FPR^{-1}(x)) dx$$

Gdzie TPR oznacza czulość, a FPR oznacza 1 – specyficzność.

W ramach przeprowadzonych analiz w pracach zamieszczano również macierze błędów, które stanowią istotne narzędzie wizualizacji wyników klasyfikacji. Macierz błędów zawiera informacje o liczbie prawdziwie pozytywnych (TP), prawdziwie negatywnych (TN), fałszywie pozytywnych (FP) oraz fałszywie negatywnych (FN) klasyfikacji, co

pozwała na szczegółową ocenę działania modelu. Dzięki temu możliwe było przejrzyste przedstawienie wyników klasyfikacji, a także identyfikacja potencjalnych obszarów, w których model mógłby wymagać poprawy. Macierze błędów umożliwiają swobodną analizę wyników, zarówno w kontekście ogólnej skuteczności modelu, jak i jego wydajności w odniesieniu do poszczególnych klas. Ich zastosowanie pozwala na łatwe zidentyfikowanie przypadków, w których model popełnia błędy, co jest szczególnie istotne w przypadku niezbalansowanych zbiorów danych. Dzięki temu narzędziu możliwe było nie tylko porównanie modeli, ale również lepsze zrozumienie ich działania w praktycznych zastosowaniach.

Tabela 1. Macierz błędów

	Przewidywana		
Rzeczywista		Negatywna	Pozytywna
	Negatywna	TN	FP
	Pozytywna	FN	TP

3.1.3. Wartości Shapley’a

W przeprowadzonych analizach wykorzystano wartości Shapley’a (SHAP), które stanowią jedno z najpopularniejszych narzędzi do interpretacji wyników modeli uczenia maszynowego^{39,40}. Wartości SHAP opierają się na teorii gier kooperacyjnych i pozwalają na wyjaśnienie wpływu poszczególnych cech na predykcję modelu. Każdej cesze przypisywana jest wartość, która reprezentuje jej wkład w końcowy wynik predykcji, co umożliwia zarówno globalną, jak i indywidualną interpretację działania modelu.

Obliczanie wartości SHAP polega na analizie wpływu danej cechy na wynik modelu poprzez porównanie predykcji z tą cechą i bez niej, uwzględniając wszystkie możliwe kombinacje cech. Dzięki temu metoda

ta zapewnia sprawiedliwą i zrównoważoną ocenę znaczenia cech, niezależnie od ich kolejności w modelu.

Równanie 6. Wartość Shapley'a

$$f(x_i) = \phi_0 + \sum_{j=1}^m \phi_{ij}$$

Gdzie ϕ_0 oznacza predykcję modelu bazowego (tj. średnią wartość wyjściową), a ϕ_{ij} mierzy addytywny wpływ cechy j na predykcję $f(x_i)$.

Globalne wyjaśnienia wyników pozwalają na ocenę ogólnego wpływu poszczególnych cech na predykcje w całym zbiorze danych, co umożliwia identyfikację najważniejszych parametrów w kontekście analizowanego problemu. Z kolei indywidualne wyjaśnienia wyników dostarczają informacji o tym, w jaki sposób konkretne cechy wpłynęły na predykcję dla pojedynczego przypadku, co jest szczególnie przydatne w praktyce klinicznej, gdzie kluczowe jest zrozumienie ryzyka dla konkretnego pacjenta. Zastosowanie wartości SHAP w pracy pozwoliło na zwiększenie interpretowalności modeli, co jest istotnym krokiem w kierunku ich praktycznego wykorzystania w środowisku klinicznym. Dzięki tej metodzie możliwe było nie tylko zidentyfikowanie kluczowych czynników ryzyka, ale również lepsze zrozumienie mechanizmów decyzyjnych modeli.

3.1.4. LASSO

W pracy zastosowano metodę LASSO (*Least Absolute Shrinkage and Selection Operator*) jako narzędzie do redukcji liczby cech w zbiorze danych. LASSO jest techniką regularyzacji, która wprowadza karę za sumę wartości bezwzględnych współczynników regresji. Dzięki temu metoda ta nie tylko zmniejsza wartości współczynników, ale również eliminuje mniej istotne cechy, przypisując ich współczynnikom wartość równą zero. W

efekcie LASSO pozwala na jednoczesną selekcję cech i dopasowanie modelu, co czyni ją szczególnie przydatną w analizach, w których liczba zmiennych jest duża w stosunku do liczby obserwacji.

Dzięki zastosowaniu LASSO możliwe było zidentyfikowanie kluczowych cech, które miały największy wpływ na wynik modelu predykcyjnego. Redukcja liczby cech pozwoliła na uproszczenie modeli, co przełożyło się na ich większą interpretowalność i efektywność obliczeniową. Jest to szczególnie istotne w analizach medycznych, gdzie kluczowe znaczenie ma identyfikacja najważniejszych czynników ryzyka. Zaletą metody LASSO w redukcji cech jest jej zdolność do eliminacji zmiennych, które w niewielkim stopniu wpływają na wynik modelu, co zmniejsza ryzyko przeuczenia i poprawia ogólną wydajność modelu. Wadą jest natomiast jej ograniczenie w przypadku silnie skorelowanych cech, gdzie metoda może arbitralnie wybrać jedną z nich, ignorując pozostałe.

Dla danego zbioru danych $X \in R^{m \times q}$ oraz wektora odpowiedzi $y \in R^m$, metoda LASSO rozwiązuje następujący problem optymalizacyjny:

Równanie 7. LASSO

$$\hat{\gamma} = \underset{\gamma}{\operatorname{arg\,min}} \left\{ \frac{1}{2m} \sum_{i=1}^m \left(y_i - \sum_{k=1}^q X_{ik} \gamma_k \right)^2 + \lambda \sum_{k=1}^q |\gamma_k| \right\}$$

Gdzie $\lambda \geq 0$ kontroluje siłę kary, zmniejszając wartości współczynników oraz eliminując nieistotne cechy.

W pracy metoda LASSO odegrała kluczową rolę w procesie selekcji cech, umożliwiając skoncentrowanie się na najbardziej istotnych zmiennych i poprawę jakości wyników modeli predykcyjnych.

3.1.5. Analiza statystyczna

W pracy przeprowadzono analizę statystyczną w celu porównania charakterystyk bazowych grup pacjentów. Zmienne katégoryczne zostały

przedstawione jako liczności i procent, a ich porównanie przeprowadzono za pomocą testu chi-kwadrat. Zmienne ciągłe zostały zaprezentowane jako średnia $\pm$ odchylenie standardowe lub jako mediana oraz pierwsze i trzecie kwartyle (Q1, Q3). Porównanie zmiennych ciągłych przeprowadzono za pomocą testu U Manna-Whitneya. Przed przystąpieniem do analizy zmiennych ciągłych sprawdzono normalność ich rozkładu za pomocą testu Shapiro-Wilka. Wyniki uznawano za statystycznie istotne, gdy wartość p była mniejsza niż 0,05. Analiza statystyczna pozwoliła na identyfikację różnic w charakterystykach wyjściowych pomiędzy grupami, co stanowiło podstawę do dalszych analiz predykcyjnych.

3.1.6. Optymalizacja hiperparametrów

W celu poprawy skuteczności opracowanych modeli predykcyjnych zastosowano optymalizację hiperparametrów, która pozwala na znalezienie najlepszych wartości parametrów sterujących działaniem algorytmów. Hiperparametry, takie jak liczba drzew, głębokość drzewa, współczynnik uczenia czy parametry regularyzacji, mają kluczowy wpływ na wydajność modeli, dlatego ich odpowiedni dobór jest niezbędny do uzyskania optymalnych wyników. W pracy wykorzystano bibliotekę Hyperopt, która umożliwia efektywną optymalizację hiperparametrów za pomocą metod przeszukiwania przestrzeni parametrów.

Hyperopt opiera się na technikach optymalizacji bayesowskiej, które pozwalają na iteracyjne przeszukiwanie przestrzeni hiperparametrów w sposób bardziej efektywny niż tradycyjne metody, takie jak przeszukiwanie siatki (*grid search*) czy losowe przeszukiwanie (*random search*)⁴¹.

Równanie 8. Optymalizacja HyperOpt

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \mathcal{L}(\theta, D)$$

gdzie:

- $\hat{\theta}$ oznacza optymalne wartości hiperparametrów,
- θ to wektor hiperparametrów podlegających optymalizacji,
- Θ reprezentuje przestrzeń dopuszczalnych wartości hiperparametrów,
- $\mathcal{L}(\theta, D)$ jest funkcją celu, np. funkcją straty lub błędu walidacyjnego, obliczaną na zbiorze danych D ,
- D to zbiór danych używany do oceny jakości rozwiązania.

Optymalizacja hiperparametrów za pomocą metody Hyperopt, w szczególności z wykorzystaniem estymatora Parzena w strukturze drzewiastej (ang. *Tree-structured Parzen Estimator*, TPE), polega na modelowaniu funkcji celu jako rozkładu prawdopodobieństwa.

W podejściu TPE, zamiast bezpośredniego przeszukiwania przestrzeni hiperparametrów, funkcja celu $\mathcal{L}(\theta, D)$ jest aproksymowana za pomocą dwóch rozkładów warunkowych:

$l(x)$: rozkład dla wartości hiperparametrów, które prowadzą do niskich wartości funkcji celu,

$g(x)$: rozkład dla pozostałych wartości hiperparametrów.

Optymalizacja polega na maksymalizacji stosunku $\frac{l(x)}{g(x)}$, co pozwala efektywnie eksplorować przestrzeń hiperparametrów, koncentrując się na obszarach o wysokim prawdopodobieństwie uzyskania lepszych wyników.

Zastosowanie Hyperopt w pracy pozwoliło na znaczącą poprawę wydajności modeli predykcyjnych, co przełożyło się na wyższe wartości metryk oceny, takich jak AUC, czułość i swoistość. Optymalizacja

hiperparametrów odegrała kluczową rolę w procesie budowy modeli, zapewniając ich lepsze dopasowanie do analizowanych danych.

3.1.7. Środowisko obliczeniowe

Wszystkie obliczenia przeprowadzono w środowisku Python 3.10.12 z wykorzystaniem ogólnodostępnych bibliotek. Obliczenia wykonano na komputerze wyposażonym w procesor z 24 wątkami o taktowaniu 4,5 GHz oraz 64 GB pamięci RAM, co zapewniło odpowiednią wydajność i szybkość przetwarzania danych. Wykorzystane wersje bibliotek obejmowały: *pandas* (2.1.4), *numpy* (1.26.4), *hyperopt* (0.2.7), *shap* (0.46.0), *scikit-learn* (1.4.2), *xgboost* (2.1.3) oraz *scipy* (1.11.4).

3.2. Publikacja I

Artificial Intelligence Technologies in Cardiology. J. Cardiovasc. Dev. Dis. 2023, 10, 202

Przegląd miał charakter narracyjny i jego celem było przedstawienie pełnego zakresu zastosowań sztucznej inteligencji w kardiologii, ze szczególnym uwzględnieniem algorytmów ML oraz technologii XAI. W ramach pracy przeanalizowano literaturę naukową dostępną w bazach danych PubMed oraz Google Scholar. Dobór literatury opierał się na subiektywnej ocenie jej znaczenia dla tematu, a kryterium włączenia stanowiły publikacje z ostatnich 10 lat, które miały istotny wpływ na rozwój AI w medycynie.

Analiza obejmowała klasyfikację algorytmów ML na trzy główne typy: uczenie nadzorowane, nienadzorowane oraz uczenie przez wzmocnienie. Szczególną uwagę poświęcono metrykom wydajności stosowanym w zadaniach regresji i klasyfikacji oraz technikom XAI. Dodatkowo, przeanalizowano przykłady zastosowań AI w kardiologii, w tym przetwarzanie języka naturalnego (NLP), oraz omówiono aspekty

prawne, etyczne i metodologiczne związane z wdrażaniem modeli AI w praktyce klinicznej.

3.3. Publikacja II

Predicting Mortality in Atrial Fibrillation Patients Treated with Direct Oral Anticoagulants: A Machine Learning Study Based on the MIMIC-IV Database. J. Clin. Med. 2025, 14, 3697

3.3.1. Źródło danych

Badanie koncentruje się na wykorzystaniu retrospektywnych danych z bazy Medical Information Mart for Intensive Care (MIMIC)-IV w wersji 3.1^{42–44}. Jest to obszerna, zanonimizowana baza danych pochodząca z Beth Israel Deaconess Medical Center w Bostonie, MA, zawierająca informacje o 364 627 pacjentach z lat 2008–2022. Dostęp do bazy MIMIC uzyskano poprzez uzyskanie akredytację za pośrednictwem platformy PhysioNet, która udostępnia bazę MIMIC.

3.3.2. Kryteria włączenia do badania

Do badania wybrano pacjentów z migotaniem przedsionków ($n = 32\,026$), zidentyfikowanych na podstawie kodów ICD ['42731', 'I4891', 'I48', 'I480', 'I481', 'I4811', 'I4819', 'I482', 'I4820', 'I4821', 'I489'], którzy rozpoczęli leczenie doustnymi antykoagulantami ($n = 6431$) (tj. dabigatran, rywaroksaban, apiksaban, edoksaban) podczas hospitalizacji. Warto zaznaczyć, że w bazie danych MIMIC-IV kody ICD występują w formie uproszczonej, bez kropek rozdzielających poszczególne segmenty kodu, co odróżnia je od standardowego zapisu stosowanego w dokumentacji klinicznej. Przykładowo, kod 'I48.91' w standardzie ICD-10 odpowiada zapisowi 'I4891' w bazie, a '427.31' (ICD-9) zapisany jest jako '42731'. Taki sposób zapisu wynika z technicznych wymogów bazy danych, które ułatwiają przetwarzanie i wyszukiwanie rekordów, eliminując znaki specjalne. Wykluczono pacjentów, którzy zmarli w

trakcie pobytu w szpitalu. Zebrano informacje dotyczące historii medycznej, pobytu w szpitalu, wskaźnika współchorobowości Charlsona (CCI), danych demograficznych (wiek, płeć, rasa, stan cywilny), pobytu w szpitalu i na oddziale intensywnej terapii, podawanych leków, wyników badań mikrobiologicznych, przeprowadzonych procedur, kodów diagnostycznych oraz wyników badań laboratoryjnych, co dało łącznie 600 zmiennych. Do analizy uwzględniono jedynie dane z pierwszej hospitalizacji, podczas której zdiagnozowano AF. Rozkład klas wynosił 6086:345. Przygotowany zbiór danych podzielono na zbiór treningowy i testowy w proporcji 80:20. Selekcję cech przeprowadzono za pomocą algorytmu LASSO, który wybrał 27 cech do budowy modelu.

3.3.3. Przygotowanie danych

Przeprowadzono czyszczenie danych, usuwając zmienne z ponad 10% brakujących wartości. W ramach inżynierii cech wyodrębniono informacje o całkowitej długości pobytu na OIT, liczbie diagnoz, długości hospitalizacji, przyczynie przyjęcia do szpitala (związanej z procedurą lub obserwacją), braku przeprowadzonych procedur, negatywnych wynikach badań mikrobiologicznych, leczeniu heparyną oraz ponownym przyjęciu w ciągu 60 dni od wypisu. Wiek pacjentów podzielono na przedziały 10-letnie i zakodowano w porządku rosnącym. Zmienne takie jak „Płeć”, „Niewydolność serca”, „Demencja”, „Cukrzyca bez powikłań”, „Paraplegia”, „Nowotwór z przerzutami”, „Ponowne przyjęcie w ciągu 60 dni”, „Stan cywilny: wdowiec”, „Przyjęcie obserwacyjne”, „Przyjęcie chirurgiczne”, „Mikrobiologia: wynik negatywny”, „Brak przeprowadzonych procedur”, „Heparyna” i „Otyłość” zakodowano jako zmienne binarne (0/1).

Pozostałe zmienne były zmiennymi ciągłymi: „Luka anionowa”, „Dwuwęglany”, „Hemoglobina”, „Potas”, „Sód”, „Azot mocznikowy”, „Białe krwinki”, „Kreatynina”, „Suma długości pobytu na oddziale intensywnej opieki”, „Liczba diagnoz”, „Wskaźnik współchorobowości Charlsona” oraz „Dni hospitalizacji”. Wszystkie zmienne ciągłe zostały standaryzowane przed zastosowaniem regresji logistycznej (LR) i algorytmu LASSO.

3.3.4. Braki danych

Rekordy pacjentów zawierające brakujące dane zostały wykluczone z analizy. Brakujące dane stanowią powszechne wyzwanie w zbiorach klinicznych, które mogą wpływać na wiarygodność i zdolność do generalizacji modeli predykcyjnych. W badaniu zastosowano wielokrotne imputacje za pomocą metody MICE, imputacje w oparciu o średnią, medianę i k-najbliższych sąsiadów oraz usuwanie listowe. Jednakże zastosowanie imputacji nie przyniosło znaczącej poprawy wyników modeli ani nie wpłynęło na główne wnioski, dlatego zdecydowano się na usuwanie listowe, aby zapewnić przejrzystość metodologiczną i interpretowalność wyników, minimalizując ryzyko wprowadzenia błędów związanych z imputacją.

3.3.5. Budowa modeli uczenia maszynowego

W celu przewidywania śmiertelności pacjentów z dowolnej przyczyny w ciągu sześciu miesięcy od wypisu ze szpitala opracowano modele uczenia maszynowego, wykorzystując algorytmy XGBoost, AdaBoost, LightGBM, RF oraz LR. Regresja logistyczna to metoda statystyczna modelująca prawdopodobieństwo wyniku binarnego za pomocą funkcji logistycznej, ceniona za prostotę i interpretowalność w zadaniach klasyfikacji binarnej.

Podczas treningu przeprowadzono wewnętrzną walidację za pomocą RepeatedStratifiedKFold ($n_splits = 5$, $n_repeats = 5$), aby zapewnić dokładność i solidność modelu. Strojenie hiperparametrów dla wszystkich modeli przeprowadzono za pomocą biblioteki HyperOpt. Dwa najlepsze modele o najwyższej wydajności w walidacji wewnętrznej wykorzystano do predykcji na zbiorze testowym.

3.4. Publikacja III

Personalized Medicine in Pulmonary Arterial Hypertension: Utilizing Artificial Intelligence for Death Prevention. J. Clin. Med. 2025, 14, 8325

3.4.1. Źródło danych

Badanie opierało się na danych z rejestru Ogólnopolskiej Bazy Nadciśnienia Płucnego (BNP-PL) ⁴⁵, który zawiera wyniki badań pacjentów z PAH zebrane z ośrodków medycznych w całej Polsce. BNP-PL jest pierwszym wieloośrodkowym, prospektywnym rejestrze pacjentów dorosłych i pediatrycznych ⁴⁶ z PAH oraz pacjentów z przewlekłym zakrzepowo-zatorowym nadciśnieniem płucnym (CTEPH) ^{47,48} w Europie Środkowo-Wschodniej ⁴⁹.

3.4.2. Kryteria włączenia do badania

Z rejestru BNP-PL wybrano dane 1755 dorosłych pacjentów z PAH. Na podstawie wizyty bazowej zebrano informacje dotyczące wzrostu, masy ciała, ciśnienia tętniczego oraz danych demograficznych. Dodatkowo uwzględniono szczegółowe wyniki badań laboratoryjnych, cewnikowania prawego serca, testów funkcjonalnych oraz echokardiografii. Ważnym elementem była historia medyczna pacjentów, obejmująca współistniejące choroby, stosowane terapie oraz przebyte procedury medyczne. Z części rejestru dotyczącej obserwacji zebrano informacje o zgonach wynikających z nagłej śmierci sercowej, niewydolności oddechowej, ostrej zatorowości płucnej, ostrego zespołu wieńcowego oraz progresji

niewydolności serca, co dało łącznie 165 zmiennych. Rozkład klas wynosił 1629:126. Przygotowany zbiór danych podzielono na zbiór treningowy i testowy w proporcji 80:20.

Selekcję cech przeprowadzono w dwóch etapach. Najpierw zastosowano algorytm LASSO, który wyeliminował 140 cech, przypisując ich współczynniki do zera. Następnie wykorzystano metodę sekwencyjnej eliminacji cech (SFE), która usunęła kolejne 8 cech, oceniając ich wpływ na dokładność klasyfikacji. Takie podejście miało na celu zwiększenie generalizację i interpretowalności modelu poprzez zachowanie jedynie najbardziej istotnych predyktorów.

3.4.3. Przygotowanie danych

Dane zostały wstępnie oczyszczone poprzez usunięcie skrajnych wartości odstających oraz duplikatów. W ramach inżynierii cech obliczono wskaźnik objętości wyrzutowej (stroke volume index) oraz stosunek amplitudy skurczowego przemieszczania pierścienia zastawki trójdzielnej (TAPSE) do skurczowego ciśnienia w tętnicy płucnej (sPAP). Liczbę stosowanych diuretyków zakodowano jako zmienną liczbową. Wszystkie zmienne kategoryczne zakodowano za pomocą metody one-hot encoding, z wyjątkiem zmiennej „Stosowanie diuretyków”, której wartość odzwierciedlała liczbę różnych leków moczopędnych stosowanych jednocześnie. W populacji BNP-PL stosowano takie diuretyki jak furosemid, torasemid, diuretyki tiazydowe oraz antagoniści aldosteronu. Wiek pacjentów podzielono na przedziały 5-letnie i zakodowano porządkowo. Zmienną docelową, czyli śmiertelność, zakodowano binarnie.

3.4.4. Braki w danych

Cechy zawierające znaczące ilości brakujących dane zostały wykluczone z analizy. Zdecydowano się na usuwanie listowe, aby

zapewnić przejrzystość metodologiczną i interpretowalność wyników, minimalizując ryzyko wprowadzenia błędów związanych z imputacją. Braki danych wynikają z faktu, że baza tworzona jest na bieżąco w różnych ośrodkach, gdzie nie wszystkie informacje nie były lub nie są dostępne w każdym z nich.

3.4.5. Budowa modeli uczenia maszynowego

Do przewidywania śmiertelności pacjentów w okresie od rozpoznania do następnej wizyty, z przyczyn sercowo-naczyniowych, opracowano model uczenia maszynowego, wykorzystując algorytm XGBoost. Algorytm ten został wybrany ze względu na wysoką wydajność, zdolność do obsługi brakujących danych, szybkość obliczeń oraz możliwość wyjaśniania predykcji za pomocą wartości SHAP. Proces treningu przeprowadzono z użyciem walidacji krzyżowej `RepeatedStratifiedKFold` ($n_splits = 10$, $n_repeats = 5$), aby zapewnić dokładność i solidność wyników. Strojenie hiperparametrów przeprowadzono za pomocą biblioteki `HyperOpt`, optymalizując takie parametry jak `colsample_bytree`, `gamma`, `learning_rate`, `max_depth`, `n_estimators`, `min_child_weight`, `reg_alpha`, `reg_lambda`, `scale_pos_weight` oraz `subsample`. Ze względu na silne niezbalansowanie klas w zbiorze treningowym przetestowano różne metody nadpróbkowania, takie jak SMOTE, ADASYN oraz SMOTETomek. Jednakże metody te nie przyniosły zadowalających rezultatów, dlatego ostatecznie wykorzystano jedynie optymalizację hiperparametrów.

4. Wyniki

4.1. Publikacja I

Artificial Intelligence Technologies in Cardiology. J. Cardiovasc. Dev. Dis. 2023, 10, 202

Sztuczna inteligencja znalazła szerokie zastosowanie w kardiologii, obejmując diagnostykę, prognozowanie oraz wspomaganie terapii chorób sercowo-naczyniowych. Algorytmy uczenia maszynowego, takie jak drzewa decyzyjne, lasy losowe, maszyny wektorów nośnych (SVM) oraz głębokie sieci neuronowe, wykazały wysoką skuteczność w analizie danych medycznych, umożliwiając precyzyjne przewidywanie wyników klinicznych.

Tabela 2. Przykłady zastosowań AI/ML w kardiologii

Autor	Badanie	Algorytm	Wydajność (Dokładność/AUC/ Precyzja)
Kogan i współ.⁵⁰	A machine learning approach to identifying patients with pulmonary hypertension using real-world electronic health records	XGBoost	0.92 AUC
Mohammad i współ.⁵¹	Development and validation of an artificial neural network algorithm to predict mortality and admission to hospital for heart failure after myocardial infarction: a	ANN	0.85–0.78 AUC

	nationwide population-based study		
Moghaddasi i współ.⁵²	Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos	SVM	Dokładność 99.45%
Attia i współ.⁹	Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram	CNN	Dokładność 85.70%
Porumb i współ.⁵³	Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG	CNN, RNN, Grad-CAM	Dokładność 82.40–85.70%
Salte i współ.⁵⁴	Artificial intelligence for automatic measurement of left ventricular strain in echocardiography	ANN	Dokładność 97–98%
Kusunose i współ.⁵⁵	A deep learning approach for assessment of regional wall motion abnormality from	CNN	0.99–0.97 AUC

	echocardiographic images		
Berikol i współ. ⁵⁶	Diagnosis of acute coronary syndrome with a support vector machine	SVM, ANN, NB, LR	Dokładność 90.10– 99.13%
Motwani i współ. ⁵⁷	Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis	DT— LogitBoos t	0.79 AUC
Kakadiaris i współ. ⁵⁸	Machine learning outperforms ACC/AHA CVD risk calculator in MESA	SVM	0.92 AUC
Kanwar i współ. ⁵⁹	Risk stratification in pulmonary arterial hypertension using Bayesian analysis	Bayesian Network	0.80 AUC
Galloway i współ. ⁶⁰	Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram	CNN	0.85–0.88 AUC
Luongo i współ. ⁶¹	Machine learning enables noninvasive prediction of atrial fibrillation driver location and acute pulmonary vein ablation success	DT	Dokładność 78.26%

	using the 12-lead ECG		
Karwath i wspól. ⁶²	Redefining β -blocker response in heart failure patients with sinus rhythm and atrial fibrillation: a machine learning cluster analysis	Hierarchical clustering, Variational autoencoders (VAEs), K-means	N/A
Cikes i wspól. ⁶³	Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy	K-means, Multiple Kernel Learning	N/A
Levy i wspól. ⁶⁴	Applications of machine learning in decision analysis for dose management for dofetilide	PCA, K-means, reinforcement learning—SARSA	Dokładność 86–93%
Garvin i wspól. ⁶⁵	Automating quality measures for heart failure using natural language processing: a descriptive study in the department of veterans affairs	NLP	Precyzja 98.7%
Shah i wspól. ⁶⁶	Impact of different electronic cohort definitions to identify patients with atrial fibrillation from the electronic medical record	NLP	0.89 AUC

Kaspar i współ. ⁶⁷	Underestimated prevalence of heart failure in hospital inpatients: a comparison of ICD codes and discharge letter information	NLP	Precyzja 96%
Patel i współ. ⁶⁸	Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records	NLP	Precyzja 96%
Mahajan i współ. ⁶⁹	Combining structured and unstructured data for predicting risk of readmission for heart failure patients	NLP	0.65 AUC
Galper i współ. ⁷⁰	Comparison of adverse event and device problem rates for transcatheter aortic valve replacement and mitraclip procedures as reported by the transcatheter valve therapy registry and the Food and Drug Administration postmarket surveillance data	NLP	N/A
Afzal i współ. ⁷¹	Mining peripheral arterial disease	NLP	Dokładność 91.8%

	cases from narrative clinical notes using natural language processing		
Ashburner i wspól.⁷²	Natural language processing to improve prediction of incident atrial fibrillation using electronic health records	NLP	N/A

Technologie XAI odegrały kluczową rolę w zwiększaniu przejrzystości i interpretowalności wyników modeli AI. Dzięki nim możliwe jest lepsze zrozumienie działania algorytmów, co ma istotne znaczenie dla ich akceptacji w środowisku medycznym.

Pomimo obiecujących wyników, wdrożenie AI w medycynie wiąże się z licznymi wyzwaniami. Kluczowe ograniczenia obejmują konieczność spełnienia rygorystycznych wymagań prawnych i etycznych, a także potrzebę standaryzacji metodologii oraz walidacji modeli na różnych populacjach pacjentów. Niemniej jednak, AI ma potencjał do zrewolucjonizowania systemów opieki zdrowotnej, przyczyniając się do redukcji kosztów, przyspieszenia procesów badawczych oraz poprawy jakości opieki nad pacjentem.

4.2. Publikacja II

Predicting Mortality in Atrial Fibrillation Patients Treated with Direct Oral Anticoagulants: A Machine Learning Study Based on the MIMIC-IV Database. J. Clin. Med. 2025, 14, 3697

4.2.1. Analiza statystyczna

W Tabeli 3. zestawiono wszystkie cechy, które zostały wyselekcjonowane do budowy modeli predykcyjnych.

Tabela 3. Analiza statystyczna 27 cech użytych do budowy modelu

Cecha	Przeżycia (n = 6086)	Zgony (n = 345)	Wartość <i>p</i> (test dwustronny)
Wiek [lata], mediana (Q1, Q3)	76 (68,84)	81 (73,88)	<0.001
Płeć, mężczyźni, n (%)	3416 (56.13%)	167 (48.41%)	0.006
Niewydolność serca, n (%)	2768 (45.48%)	243 (70.43%)	<0.001
Demencja, n (%)	398 (6.54%)	49 (14.20%)	<0.001
Cukrzyca bez powikłań, n (%)	1313 (21.57%)	77 (22.32%)	0.795
Paraplegia, n (%)	316 (5.19%)	54 (15.65%)	<0.001
Nowotwór z przerzutami, n (%)	192 (3.15%)	39 (11.30%)	<0.001
Ponowne przyjęcie w ciągu 60 dni, n (%)	1249 (20.52%)	120 (34.78%)	<0.001
Stan cywilny: Wdowiec, n (%)	1051 (17.27%)	74 (21.45%)	0.055
Przyjęcie obserwacyjne, n (%)	2702 (44.40%)	103 (29.86%)	<0.001
Przyjęcie chirurgiczne, n (%)	418 (6.87%)	5 (1.45%)	<0.001
Mikrobiologia: wynik negatywny, n (%)	5113 (84.01%)	191 (55.36%)	<0.001
Brak przeprowadzonych procedur, n (%)	2197 (36.10%)	53 (15.36%)	<0.001
Heparyna, n (%)	3984 (65.46%)	313 (90.72%)	<0.001

Otyłość, n (%)	831 (13.65%)	37 (10.72%)	0.142
Luka anionowa [mEq/L], średnia ± SD	15.89 ± 3.55	18.33 ± 4.27	<0.001
Dwuwęglany [mEq/L], średnia ± SD	27.79 ± 4.04	30.58 ± 5.30	<0.001
Kreatynina [mg/dL], średnia ± SD	1.43 ± 1.28	2.05 ± 1.87	<0.001
Hemoglobina [g/dL], średnia ± SD	12.45 ± 2.07	11.6 ± 2.08	<0.001
Potas [mEq/L], średnia ± SD	4.69 ± 0.71	5.26 ± 0.98	<0.001
Sód [mEq/L], średnia ± SD	141.88 ± 4.01	144.94 ± 5.67	<0.001
Azot mocznikowy [mg/dL], średnia ± SD	30.76 ± 21.11	48.97 ± 27.20	<0.001
Białe krwinki [10⁹/μL], średnia ± SD	12.3 ± 9.55	17.35 ± 10.19	<0.001
Suma długości pobytu na oddziale intensywnej opieki [dni], mediana (Q1, Q3)	0 (0, 1.45)	3.19 (1.68, 7.73)	<0.001
Liczba diagnoz, mediana (Q1, Q3)	17 (12,22)	27 (21,33)	<0.001
Wskaźnik współchorobowości Charlsona, mediana (Q1, Q3)	5 (4,7)	8 (6,10)	<0.001
Dni hospitalizacji [dni], mediana (Q1, Q3)	5 (2,9)	13 (8,23)	<0.001

Przeprowadzona analiza danych wskazuje na istotne różnice między pacjentami, którzy przeżyli (n = 6086), a tymi, którzy zmarli (n = 345). Mediana wieku w grupie osób zgonów była wyższa (81 lat, Q1 = 73, Q3 = 88) w porównaniu do grupy przeżywających (76 lat, Q1 = 68, Q3 = 84). W grupie zgonów odsetek mężczyzn był niższy (48.41%) niż w grupie przeżywających (56.13%). Występowanie niewydolności serca było znacznie częstsze w grupie zgonów (70.43%) niż w grupie przeżywających

(45.48%). Podobnie, demencja (14.20% vs. 6.54%), paraplegia (15.65% vs. 5.19%) oraz nowotwór z przerzutami (11.30% vs. 3.15%) były bardziej powszechne w grupie zgonów.

Ponowne przyjęcie w ciągu 60 dni było częstsze w grupie zgonów (34.78%) niż w grupie przeżywających (20.52%). Przyjęcia obserwacyjne były bardziej typowe dla grupy przeżywających (44.40% vs. 29.86%), podczas gdy przyjęcia chirurgiczne były rzadsze w grupie zgonów (1.45% vs. 6.87%). Negatywne wyniki badań mikrobiologicznych były częstsze w grupie przeżywających (84.01% vs. 55.36%), a brak przeprowadzonych procedur był bardziej powszechny w tej grupie (36.10% vs. 15.36%). Stosowanie heparyny było częstsze w grupie zgonów (90.72% vs. 65.46%).

Analiza zmiennych ciągłych wykazała, że pacjenci z grupy zgonów mieli wyższe średnie wartości luki anionowej (18.33 ± 4.27 vs. 15.89 ± 3.55 mEq/L), dwuwęglanów (30.58 ± 5.30 vs. 27.79 ± 4.04 mEq/L), kreatyniny (2.05 ± 1.87 vs. 1.43 ± 1.28 mg/dL), potasu (5.26 ± 0.98 vs. 4.69 ± 0.71 mEq/L;), sodu (144.94 ± 5.67 vs. 141.88 ± 4.01 mEq/L), azotu mocznikowego (48.97 ± 27.20 vs. 30.76 ± 21.11 mg/dL) oraz białych krwinek (17.35 ± 10.19 vs. $12.3 \pm 9.55 \times 10^9/\mu\text{L}$). Z kolei średnia hemoglobina była niższa w grupie zgonów (11.6 ± 2.08 vs. 12.45 ± 2.07 g/dL).

Mediana długości pobytu na oddziale intensywnej opieki była znacznie wyższa w grupie zgonów (3.19 dni, Q1 = 1.68, Q3 = 7.73) w porównaniu do grupy przeżywających (0 dni, Q1 = 0, Q3 = 1.45). Liczba diagnoz była również wyższa w grupie zgonów (mediana = 27, Q1 = 21, Q3 = 33) niż w grupie przeżywających (mediana = 17, Q1 = 12, Q3 = 22). Wskaźnik współchorobowości Charlsona był wyższy w grupie zgonów (mediana = 8, Q1 = 6, Q3 = 10) w porównaniu do grupy przeżywających

(mediana = 5, Q1 = 4, Q3 = 7). Dni hospitalizacji były dłuższe w grupie zgonów (mediana = 13, Q1 = 8, Q3 = 23) niż w grupie przeżywających (mediana = 5, Q1 = 2, Q3 = 9).

Podsumowując, grupa pacjentów, którzy zmarli, charakteryzowała się starszym wiekiem, większym obciążeniem współistniejącymi chorobami oraz gorszymi wynikami klinicznymi w porównaniu do grupy przeżywających. Brak różnic pomiędzy grupami pacjentów, przy założeniu istotności statystycznej $p < 0.05$, odnotowano dla 3 cech: „Cukrzyca bez powikłań”; „Stan cywilny: wdowiec”; „Otyłość” (odpowiednio $p = 0.795$; $p = 0.055$; $p = 0.142$).

4.2.2. Metryki

W przeprowadzonych analizach porównano wyniki pięciu algorytmów uczenia maszynowego, które zostały ocenione na podstawie wskaźników AUC, MCC, czułości, specyficzności oraz dokładności. Kluczowymi czynnikami decydującymi o wyborze finalnego algorytmu były wskaźniki MCC oraz AUC, które skutecznie oceniają jakość modelu klasyfikacyjnego w kontekście silnie niezrównoważonego zbioru danych. W Tabeli 4 przedstawiono podsumowanie wszystkich metryk dla każdego modelu, wraz z 95% przedziałami ufności (CI) dla wyników uzyskanych podczas wewnętrznej walidacji krzyżowej.

Tabela 4. Metryki wydajności na zbiorze treningowym

Algorytm	AUC (95%CI)	Dokładność (95%CI)	Czułość (95%CI)	Specyficzność (95%CI)	MCC (95%CI)
LightGBM	0.873 (0.865– 0.881)	0.862 (0.856– 0.867)	0.886 (0.870– 0.901)	0.860 (0.854– 0.866)	0.439 (0.427– 0.450)

XGBoost	0.871 (0.863– 0.879)	0.846 (0.840– 0.853)	0.899 (0.882– 0.915)	0.843 (0.836– 0.850)	0.422 (0.411– 0.433)
RF	0.862 (0.854– 0.870)	0.853 (0.846– 0.859)	0.873 (0.855– 0.890)	0.851 (0.845– 0.858)	0.419 (0.409– 0.429)
AdaBoost	0.844 (0.839– 0.849)	0.764 (0.745– 0.783)	0.934 (0.908– 0.961)	0.754 (0.733– 0.775)	0.348 (0.340– 0.357)
LR	0.795 (0.785– 0.805)	0.770 (0.765– 0.776)	0.822 (0.801– 0.842)	0.767 (0.762– 0.773)	0.301 (0.291– 0.312)

Najlepsze wyniki osiągnęły algorytmy LightGBM oraz XGBoost, przy czym LightGBM wykazał się wyższymi wartościami wszystkich metryk (Tabela 5), ze szczególnym uwzględnieniem wysokiej czułości modelu. Wysoka czułość jest kluczowa w kontekście przewidywania śmiertelności pacjentów, ponieważ pozwala na skuteczniejsze identyfikowanie przypadków pozytywnych. Jednocześnie model LightGBM osiągnął satysfakcjonujące wyniki specyficzności, co świadczy o jego zdolności do minimalizowania liczby fałszywie pozytywnych klasyfikacji. Test Z przeprowadzony na wynikach uzyskanych przez modele LightGBM i XGBoost na danych testowych nie wykazał istotnych statystycznie różnic w czułości ani specyficzności ($p \gg 0,05$).

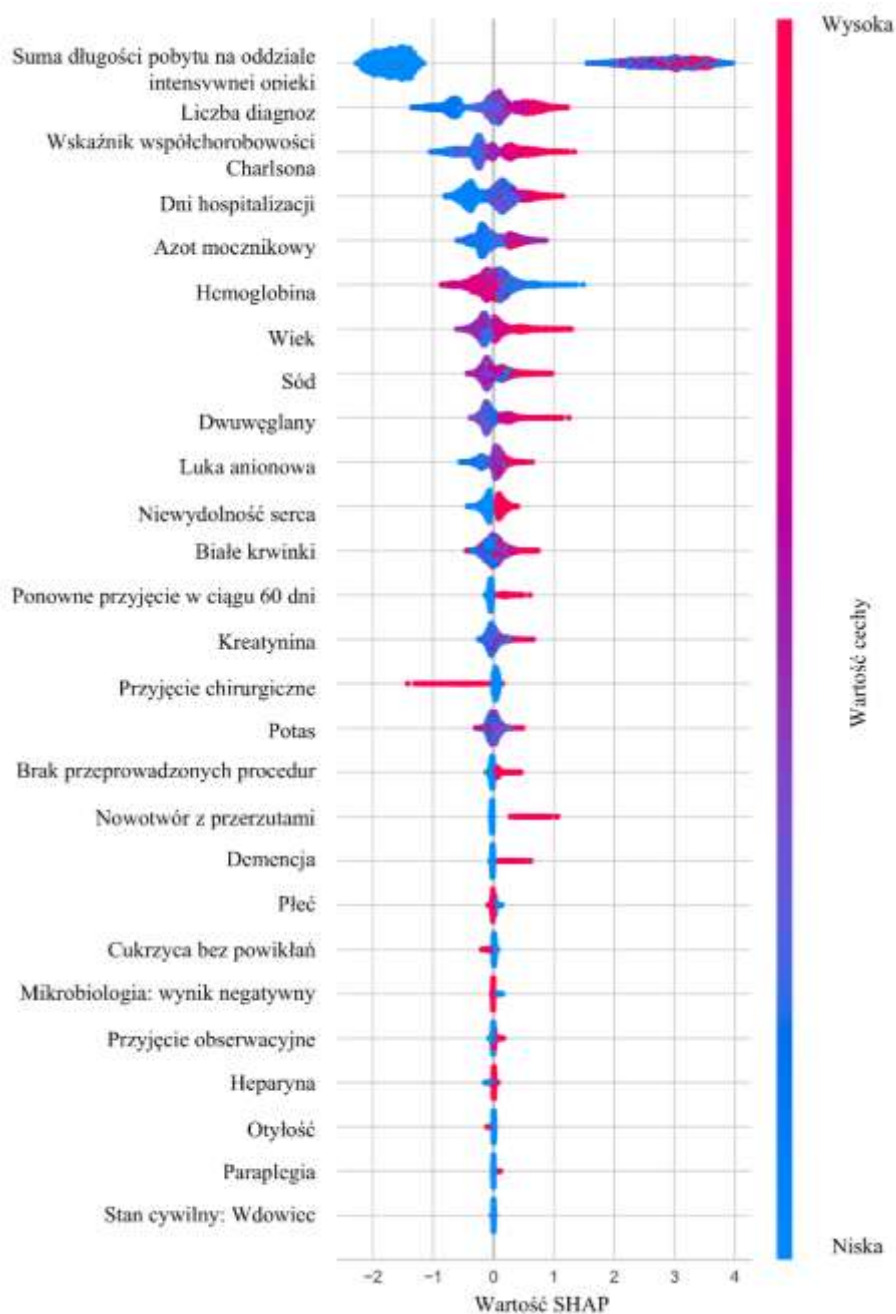
Tabela 5. Metryki wydajności na zbiorze testowym

Algorytm	AUC (95%CI)	Dokładność (95%CI)	Czułość (95%CI)	Specyficzność (95%CI)	MCC (95%CI)
LightGBM	0.886 (0.849– 0.919)	0.862 (0.842– 0.880)	0.913 (0.842– 0.973)	0.859 (0.839– 0.878)	0.450 (0.388– 0.507)

XGBoost	0.873 (0.834– 0.908)	0.849 (0.828– 0.868)	0.899 (0.823– 0.967)	0.846 (0.825– 0.866)	0.425 (0.362– 0.483)
---------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------

4.2.3. SHAP

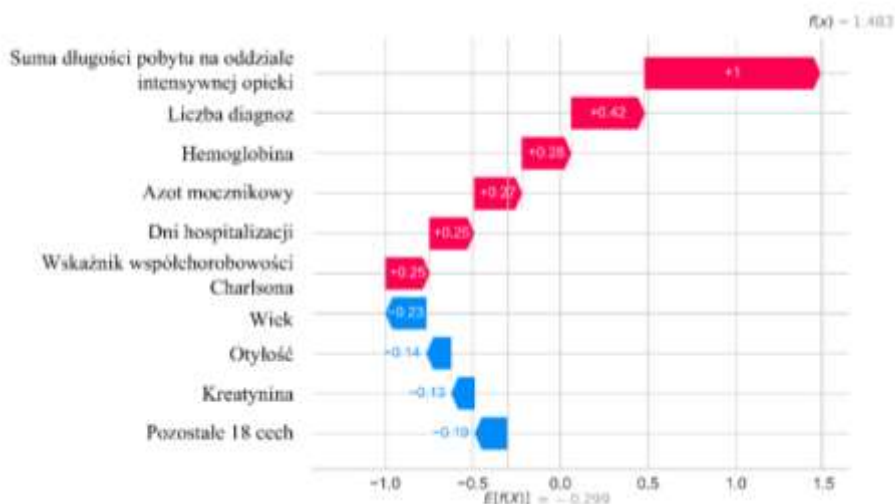
Do wyjaśnienia uzyskanych wyników wykorzystano SHAP, za pomocą, której wskazano cechy mające największy wpływ na predykcję w ujęciu globalnym i lokalnym. Na Rycinie 4.2.3 przedstawiono wykres wizualizujący wpływ poszczególnych cech na predykcję modelu za pomocą wartości SHAP. Pozycja na osi odciętych wskazuje wielkość i kierunek wpływu danej cechy na predykcję, natomiast kolor reprezentuje wartość cechy. Cechy uszeregowane są od mającej największy wpływ na zdolność predykcyjną do mającej najmniejszy wpływ. Najistotniejsze cechy obejmują sumę długości pobytu na oddziale intensywnej opieki medycznej, wskaźnika współchorobowości Charlsona i liczby diagnoz oraz liczby dni hospitalizacji. Spośród wyników laboratoryjnych największą wagę model przypisał dla poziomu azotu mocznikowego, hemoglobiny, sodu oraz dwuwęglanów.



Rycina 1. Wykres SHAP globalny

Przykładowy wykres interpretacji lokalnej przedstawiono na Rycina 2. Bez możliwości prześledzenia procesu decyzyjnego modelu trudno jest

zdołyć zaufanie klinicystów. Wartości SHAP są w tym kontekście niezwykle skuteczne. Lokalne wyjaśnienia dla poszczególnych pacjentów, wskazujące, co napędza predykcje modelu, nie tylko zwiększają zaufanie do uzyskanych wyników, ale także umożliwiają podejście spersonalizowane.



Rycina 2. Wykres SHAP lokalny dla pacjenta, u którego model prawidłowo przewidział zgon.

4.3. Publikacja III

Personalized Medicine in Pulmonary Arterial Hypertension: Utilizing Artificial Intelligence for Death Prevention. J. Clin. Med. 2025, 14, 8325

4.3.1. Analiza statystyczna

W Tabeli 6 zestawiono cechy, które zostały wyselekcjonowane do budowy modelu predykcyjnego.

W okresie od 1 marca 2018 roku do 31 sierpnia 2023 roku do rejestru włączono łącznie 1755 pacjentów z PAH, w tym 902 nowo zdiagnozowanych oraz 853 wcześniej zdiagnozowanych (przed 1 marca 2018 roku). W tym samym czasie zmarło 126 pacjentów. Mediana i średni czas od wizyty bazowej do zgonu wynosiły odpowiednio 10 i 12 miesięcy.

Kohorta badawcza obejmowała zróżnicowany zakres etiologii PAH. Najczęstszym typem było idiopatyczne, niereaktywne PAH (750 pacjentów, 8% zgonów), następnie PAH związane z wrodzonymi wadami serca (483 pacjentów, 5% zgonów) oraz PAH związane z chorobami tkanki łącznej (305 pacjentów, 10% zgonów). Rzadziej występowały takie typy jak idiopatyczne, reaktywne PAH (121 pacjentów, 2% zgonów), PAH związane z nadciśnieniem wrotnym (45 pacjentów, 7% zgonów), dziedziczne PAH (19 pacjentów, 5% zgonów), PAH związane z zakażeniem HIV (10 pacjentów, 0% zgonów) oraz PAH wywołane lekami lub toksynami (14 pacjentów, 14% zgonów). Choroba żylnych okluzyjnych naczyń płucnych była rzadka (6 pacjentów), ale charakteryzowała się najwyższym wskaźnikiem śmiertelności (33%). Tylko dwa przypadki zostały sklasyfikowane jako PAH o nieznannej etiologii. Charakterystyka bazowa pacjentów z PAH została porównana w odniesieniu do śmiertelności wynikającej z nagłej śmierci sercowej, niewydolności oddechowej, ostrej zatorowości płucnej, ostrego zespołu wieńcowego oraz progresji niewydolności serca w populacji pacjentów z PAH.

Przeprowadzona analiza danych wykazała istotne różnice między pacjentami, którzy przeżyli ($n = 1629$), a tymi, którzy zmarli ($n = 126$). Mediana wieku w grupie osób zgonów była wyższa (68 lat, $Q1 = 57.75$, $Q3 = 74$) w porównaniu do grupy przeżywających (59 lat, $Q1 = 40$, $Q3 = 70$). W grupie zgonów odsetek mężczyzn był wyższy (40.48%) niż w grupie przeżywających (31.25%).

Występowanie przewlekłej choroby nerek było znacznie częstsze w grupie zgonów (35.71%) niż w grupie przeżywających (16.82%). Podobnie, cukrzyca (35.71% vs. 19.77%) oraz przebyty zawał mięśnia

sercowego (16.67% vs. 4.79%) były bardziej powszechne w grupie zgonów. Hospitalizacje z powodu PAH w ciągu ostatniego roku były częstsze w grupie zgonów (31.36%) niż w grupie przeżywających (15.81%). Historia przedłużonych hospitalizacji również była bardziej powszechna w grupie zgonów (16.67% vs. 8.59%).

Objawy omdleń w ciągu ostatniego roku były częstsze w grupie zgonów (15.25%) niż w grupie przeżywających (6.39%). Objawy niewydolności prawej komory serca występowały znacznie częściej w grupie zgonów (63.56%) niż w grupie przeżywających (32.48%). Odchylenie osi serca w EKG było bardziej powszechne w grupie zgonów (54.76%) niż w grupie przeżywających (38.43%), podobnie jak obecność płynu w worku osierdziowym (33.33% vs. 18.07%).

Wartość wskaźnika TAPSE/sPAP była niższa w grupie zgonów (mediana = 0.21, Q1 = 0.15, Q3 = 0.28) w porównaniu do grupy przeżywających (mediana = 0.25, Q1 = 0.18, Q3 = 0.38). Dystans w 6-minutowym teście marszowym był znacznie krótszy w grupie zgonów (mediana = 231 m, Q1 = 120, Q3 = 322.5) niż w grupie przeżywających (mediana = 360 m, Q1 = 240, Q3 = 460). Stężenie NT-proBNP w osoczu było wyższe w grupie zgonów (mediana = 2983.5 pg/mL, Q1 = 1445.75, Q3 = 5529.75) w porównaniu do grupy przeżywających (mediana = 813.9 pg/mL, Q1 = 250.55, Q3 = 2482.5).

Stosowanie diuretyków było bardziej powszechne w grupie zgonów, gdzie większy odsetek pacjentów stosował 2 lub więcej rodzajów diuretyków (42.06% vs. 28.3%). Parenteralne stosowanie treprostynilu lub epoprostenolu było również częstsze w grupie zgonów (38.89%) niż w grupie przeżywających (17.25%).

Podsumowując, pacjenci, którzy zmarli, charakteryzowali się starszym wiekiem, większym obciążeniem współistniejącymi chorobami, gorszymi wynikami klinicznymi oraz wyższymi wartościami biomarkerów wskazujących na zaawansowaną niewydolność serca w porównaniu do pacjentów, którzy przeżyli. Brak różnic pomiędzy grupami pacjentów, przy założeniu istotności statystycznej $p < 0.05$, odnotowano jedynie dla „Typ PAH: związany z wrodzoną wadą serca” ($p = 0.057$).

Tabela 6. Analiza statystyczna 17 cech użytych do budowy modelu

Cecha	Przeżycia (n = 1629)	Zgony (n = 126)	Wartość p (test dwustronny)
Wiek [lata], mediana (Q1, Q3)	59 (40, 70)	68 (57.75, 74)	<0.001
Płeć, mężczyźni, n (%)	509 (31.25%)	51 (40.48%)	0.041
Typ PAH: związany z wrodzoną wadą serca, n (%)	485 (28.12%)	25 (19.84%)	0.057
Przewlekła choroba nerek, n (%)	274 (16.82%)	45 (35.71%)	<0.001
Cukrzyca, n (%)	322 (19.77%)	45 (35.71%)	<0.001
Przebyty zawał mięśnia sercowego, n (%)	78 (4.79%)	21 (16.67%)	<0.001
Hospitalizacja z powodu PAH w ostatnim roku, n (%)	240 (15.81%)	37 (31.36%)	<0.001

Historia przedłużonych hospitalizacji, n (%)	140 (8.59%)	21 (16.67%)	0.004
Omdlenia w ostatnim roku, n (%)	97 (6.39%)	18 (15.25%)	0.001
Objawy niewydolności prawej komory serca, n (%)	493 (32.48%)	75 (63.56%)	<0.001
Odchylenie osi serca w EKG, n (%)	626 (38.43%)	69 (54.76%)	<0.001
Obecność płynu w worku osierdziowym, n (%)	290 (18.07%)	39 (33.33%)	<0.001
Wskaźnik TAPSE/sPAP [mm/mmHg], mediana (Q1, Q3)	0.25 (0.18, 0.38)	0.21 (0.15, 0.28)	0.001
Dystans w 6-minutowym teście marszowym [m], mediana (Q1, Q3)	360 (240, 460)	231 (120, 322.5)	<0.001
Stężenie NT-proBNP w osoczu [pg/mL], mediana (Q1, Q3)	813.9 (250.55, 2482.5)	2983.5 (1445.75, 5529.75)	<0.001
Stosowanie diuretyków [0/1/2/3/4], n (%)	487 (29.9%) 563 (34.56%) 461 (28.3%) 109 (6.69%) 9 (0.55%)	11 (8.73%) 46 (36.51%) 53 (42.06%) 12 (9.52%) 4 (3.17%)	<0.001

Parenteralne stosowanie treprostynilu lub epoprostenolu, n (%)	281 (17.25%)	49 (38.89%)	<0.001
---	--------------	-------------	--------

4.3.2. Metryki

W przeprowadzonych analizach oceniono wydajność modelu klasyfikacyjnego opartego na algorytmie XGBoost na zbiorze testowym. Model wykazał umiarkowaną skuteczność, osiągając wskaźnik AUC na poziomie 0.767 (95% CI: 0.682–0.843) oraz dokładność wynoszącą 0.738 (95% CI: 0.695–0.783). Szczególną uwagę należy zwrócić na wysoką czułość modelu (0.800, 95% CI: 0.636–0.947), która pozwala na skuteczne identyfikowanie przypadków pozytywnych, co jest kluczowe w kontekście przewidywania śmiertelności pacjentów z PAH. Wszystkie metryki wydajności zostały przedstawione w Tabeli 7. Optymalizacja parametrów algorytmu XGBoost umożliwiła osiągnięcie tej wysokiej czułości, pomimo znacznego niezbalansowania klas w zbiorze danych.

Tabela 7. Metryki wydajności na zbiorze testowym

Metryki	Wynik	95% CI
Dokładność	0.738	0.695–0.783
Czułość	0.800	0.636–0.947
Specyficzność	0.733	0.688–0.779
MCC	0.298	0.195–0.399
AUC	0.767	0.682–0.843

Jednakże model charakteryzował się stosunkowo wysokim odsetkiem fałszywie pozytywnych predykcji, co przełożyło się na niską precyzję i niższą specyficzność (0.733, 95% CI: 0.688–0.779). Wskaźnik

MCC, wynoszący 0.298 (95% CI: 0.195–0.399), wskazuje na umiarkowaną zdolność modelu do rozróżniania między klasami.

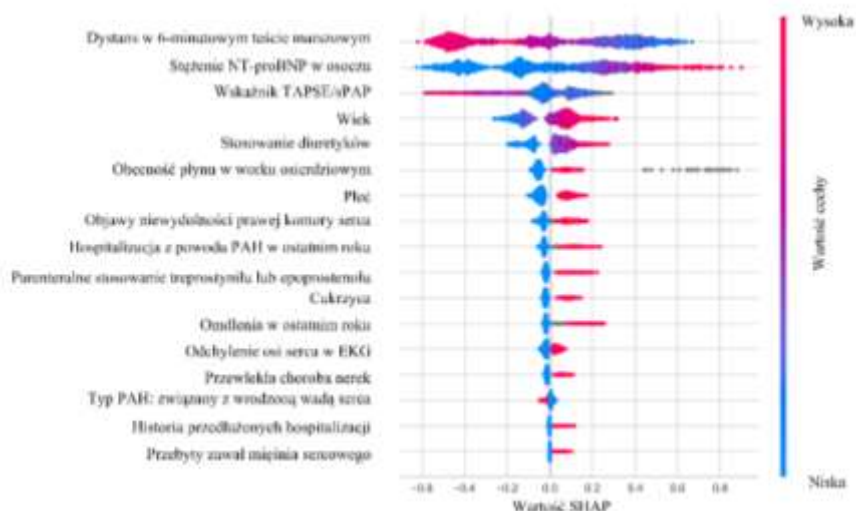
Macierz pomyłek - Tabela 8 - pokazuje, że model poprawnie sklasyfikował 239 przypadków jako negatywne oraz 20 przypadków jako pozytywne, przy czym odnotowano 87 fałszywie pozytywnych i 5 fałszywie negatywnych predykcji. Wysoka czułość modelu, przy jednoczesnym zachowaniu umiarkowanej specyficzności, wskazuje na jego przydatność w zadaniach, gdzie minimalizacja fałszywie negatywnych wyników jest priorytetem, takich jak przewidywanie śmiertelności w populacji pacjentów z PAH.

Tabela 8. Macierz błędów dla zbioru testowego

Rzeczywista	Przewidywana		
		Negatywna	Pozytywna
	Negatywna	239	87
	Pozytywna	5	20

4.3.3. SHAP

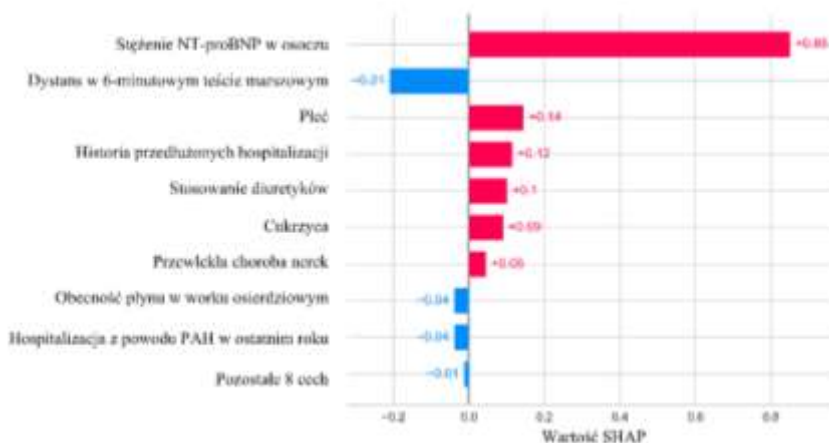
Wartości SHAP pozwoliły na identyfikację istotnych parametrów klinicznych, które mogą być wykorzystane do oceny ryzyka oraz modyfikacji leczenia pacjentów. Wykres Rycina 3 przedstawia rozkład wartości SHAP dla każdej cechy, ukazując zmienność wpływu cech na predykcje w różnych przypadkach. Wizualizacja ta pozwala zrozumieć, w jaki sposób różne wartości cech wpływają na wyniki modelu.



Rycina 3. Wykres SHAP globalny

Cechy zidentyfikowane przez model znajdują odzwierciedlenie w wytycznych ESC dotyczących stratyfikacji pacjentów do grup ryzyka, obejmując takie parametry jak stężenie NT-proBNP w osoczu, wynik 6-minutowego testu marszowego, stosunek TAPSE do sPAP oraz obecność płynu w worku osierdziowym. Wszystkie te parametry zostały uznane za mające największy wpływ na zdolność predykcyjną modelu, co potwierdzają wartości SHAP. Lokalna interpretacja najważniejszych cech

umożliwia łatwą interpretację wyników oraz potencjalną interwencję medyczną Rycina 4.



Rycina 4. Wykres SHAP lokalny dla pacjenta, u którego model prawidłowo przewidział zgon.

Wykres przedstawia dziewięć najważniejszych cech, ilustruje wartości SHAP dla konkretnego pacjenta, wskazując wkład każdej cechy w predykcję modelu. Każdy słupek reprezentuje jedną cechę, a jego długość odzwierciedla wielkość wpływu danej cechy na wynik predykcji. Dodatkowo wartości SHAP (słupki skierowane w prawo) oznaczają cechy zwiększające ryzyko zgonu, natomiast ujemne wartości SHAP (słupki skierowane w lewo) wskazują cechy zmniejszające to ryzyko.

Takie szczegółowe przedstawienie pozwala na jasne zrozumienie kluczowych czynników wpływających na decyzję modelu w przypadku tego konkretnego pacjenta. W analizowanym przypadku model przewidział, że pacjent należy do grupy zgonów, co zostało potwierdzone faktycznym przypisaniem pacjenta do tej grupy.

5. Dyskusja

Prezentowana praca miała na celu wykorzystanie metod sztucznej inteligencji do wsparcia codziennej pracy klinicznej w zakresie chorób sercowo-naczyniowych. Wykazano zadowalającą skuteczność algorytmów LightGBM i XGBoost w przewidywaniu zgonów pacjentów z AF oraz PAH. Przedstawiono użyteczność stosowania technik XAI, które w zdecydowanym stopniu podnoszą zaufanie do proponowanych rozwiązań, wskazując kluczowe czynniki wpływające na predykcję modelu.

W ostatnich latach obserwujemy dynamiczny rozwój sztucznej inteligencji w medycynie, co znajduje odzwierciedlenie zarówno w codziennej praktyce klinicznej, jak i w badaniach naukowych. Przykładem są inteligentne zegarki, które dzięki wbudowanym algorytmom AI potrafią wykrywać zaburzenia rytmu serca np. AF i wspierać pacjentów w monitorowaniu zdrowia. Jednak wdrożenie AI w medycynie wiąże się z licznymi wyzwaniami, zarówno na poziomie technicznym, prawnym i etycznym.

Na poziomie technicznym kluczowe są kwestie związane z jakością i kompletnością danych, standaryzacją procesów przetwarzania oraz przejrzystością działania stosowanych algorytmów. Modele AI wymagają dużych, reprezentatywnych i dobrze opisanych zbiorów danych, a wszelkie braki, błędy czy niejednorodność danych mogą prowadzić do błędnych predykcji i ograniczać użyteczność narzędzi w praktyce klinicznej. Dodatkowo, konieczne jest zapewnienie powtarzalności i transparentności działania algorytmów, a także ich odporności na cyberataki.

Wyzwania prawne obejmują przede wszystkim ochronę prywatności i bezpieczeństwa danych pacjentów, zgodność z przepisami takimi jak RODO, a także określenie odpowiedzialności za decyzje podejmowane na podstawie rekomendacji AI. Brak jasnych regulacji prawnych dotyczących wdrażania i stosowania AI w ochronie zdrowia może prowadzić do niepewności zarówno po stronie lekarzy, jak i pacjentów.

Z kolei wyzwania etyczne dotyczą m.in. zapewnienia sprawiedliwego i niedyskryminującego dostępu do nowych technologii, unikania uprzedzeń (bias) w modelach, a także konieczności zachowania autonomii pacjenta i roli lekarza w procesie decyzyjnym. Ważne jest, aby pacjenci i lekarze posiadali świadomość o możliwościach i ograniczeniach narzędzi AI oraz mieli prawo do świadomego wyboru i wyrażenia zgody na ich wykorzystanie w procesie leczenia.

Wszystkie te aspekty wymagają ścisłej współpracy interdyscyplinarnej, ciągłego monitorowania oraz rozwoju jasnych standardów i wytycznych, które pozwolą na bezpieczne i skuteczne wdrożenie AI w medycynie.

Zgodnie z zaleceniami ESC stosowanie DOAC w leczeniu AF jest pierwszą linią profilaktyki zakrzepowo-zatorowej ¹⁶. W licznych badaniach klinicznych i metaanalizach potwierdzono ich skuteczność i bezpieczeństwo ⁷³⁻⁷⁶. Poszukiwanie możliwości wykorzystania sztucznej inteligencji do przewidywania zgonów pacjentów z AF stało się obiektem badań w wielu ośrodkach na świecie. Chen i współ. zaproponowali nowatorskie połączenie LASSO z regresją Cox'a do przewidywania przeżycia w okresie 30, 180 i 360 dni ⁷⁷. W rezultacie otrzymali AUC = 0.826 dla przewidywania w okresie 180 dni na zbiorze treningowym i AUC = 0.788 na zbiorze walidacyjnym, co stanowi słabszy wynik niż

proponowane w pracy rozwiązanie, gdzie dla zbioru treningowego i testowego AUC wynosiło odpowiednio 0.873 i 0.886. Interesującym podejściem jest zaproponowanie nomogramu oraz interaktywnego graficznego interfejsu do obsługi modelu przez użytkownika. W pracy Lv i współ. przedstawiono rozwiązanie, w którym wykorzystano model XGBoost do przewidywania zgonów z przyczyn sercowych w trakcie pobytu w szpitalu wśród pacjentów z AF ⁷⁸. Model zbudowany na 79 cechach osiągnął AUC na zbiorze walidacyjnym na bardzo wysokim poziomie 0.932 i zaproponowano również wyjaśnienie predykcji z wykorzystaniem SHAP. Wśród najważniejszych trzech cech wskazano trijodotyroninę, tyroksynę całkowitą i wolną trijodotyroninę, których sumaryczny wpływ jest większy niż pozostałych cech. Autorzy podkreślają jednak, że wskazane przez model parametry funkcjonowania tarczycy odzwierciedlają, w większym stopniu, ogólny ciężki stan pacjenta, a nie chorobę tarczycy jako bezpośrednią przyczynę zgonu. Ze względu na ograniczenia w dostępności kompletnych danych, nie było możliwości zweryfikowania wpływu tej cech na proponowany w niniejszej rozprawie model. W pracy opublikowanej przez Nguyen i współ. zaprezentowano wyniki modelu XGBoost do przewidywania zgonów wśród pacjentów z AF hospitalizowanych na oddziale intensywnej opieki medycznej ⁷⁹. Do uczenia modelu wykorzystano bazę MIMIC – IV, a walidację przeprowadzono na bazie MIMIC – III, osiągnięto AUC na poziomie 0.743. Dodatkowo w pracy przedstawiono modele regresji do przewidywania długości pobytu na oddziale intensywnej opieki, jednak wyjaśniały niewielką część wariacji, co sugeruje, że dane z pierwszych 24 godzin hospitalizacji są niewystarczające do dokładnej predykcji tego parametru. Autorzy zaproponowali wykorzystanie SHAP do interpretacji

predykcji, wskazując na niską diurezę w ciągu pierwszych 24 godzin, niższa minimalna saturacja tlenem, wyższe maksymalne stężenie azotu mocznikowego, wyższa luka anionowa, starszy wiek, wyższa maksymalna temperatura ciała, ekspozycja na leki wazopresyjne lub inotropowe, a także charakterystyka przyjęcia, w tym przyjęcie chirurgiczne tego samego dnia i skierowanie przez lekarza. Jednak przedstawione niskie wartości precyzji i czułości modelu, odpowiednio 0.23 i 0.48, budzą wiele wątpliwości co do jakości przedstawionego rozwiązania. Niska skuteczność sugeruje konieczność dalszej optymalizacji modelu, lepszego doboru cech lub zastosowania innych algorytmów, zanim będzie można rozważać jego wdrożenie w praktyce klinicznej. W pracy opartej o dane z rejestru START-2 ⁸⁰, gdzie autorzy przedstawili różne modele do przewidywania zgonów ze wszystkich przyczyn, zgonów z przyczyn sercowo-naczyniowych, dużych krwawień i udaru w populacji pacjentów z AF stosujących leczenie przeciwkrzepliwe ⁸¹. W przedstawionym przykładzie najlepsze wyniki uzyskiwał model oparty o Multi-Gate Mixture of Experts jest to architektura modelu uczenia maszynowego, w której wiele niezależnych „ekspertów” (podmodeli) jest sterowanych przez kilka bramek (gate), które decydują, w jakim stopniu każdy ekspert wpływa na końcową predykcję. W odróżnieniu od klasycznego Mixture of Experts, gdzie jedna bramka wybiera lub waży ekspertów, w Multi-Gate Mixture of Experts każda wyjściowa ścieżka może mieć własną, niezależną bramkę, co pozwala na bardziej elastyczne i zróżnicowane łączenie wiedzy poszczególnych ekspertów. Podejście wykorzystywało osiem sieci ekspertów, z których każda miała po 16 jednostek w części współdzielonej oraz warstwę 32 wyspecjalizowanych jednostek dla każdego zadania. W obu przypadkach do uczenia modeli zastosowano optymalizator Adam

(liczba epok = 100). Dodatkowo dostrajano kluczowe hiperparametry architektur głębokich, takie jak liczba jednostek w warstwach, stałe regularyzacji L1/L2 oraz współczynniki dropout, aby uzyskać bardziej odporne modele i ograniczyć przeuczenie. Autorzy uzyskali wynik AUC dla zgonów ze wszystkich przyczyn i zgonów z przyczyn sercowo-naczyniowych w próbie pacjentów stosujących DOAC na poziomie 0.733. Autorzy dodatkowo przedstawili wyniki skal HASBLED i CHA₂DSVA₂SC porównując je do wyników uzyskanych przez modele, wykazując wyższość proponowanego rozwiązania w przewidywaniu zgonów, mając na uwadze inne założenia przygotowanych skal ryzyka i ich umiarkowane przełożenie na predykcję zgonów ¹⁹. Wadą proponowanego rozwiązania jest ograniczony wgląd w wytłumaczalność modelu oraz brak przedstawionych innych metryk poza AUC, przez co ciężko jest w pełni ocenić możliwości prezentowane w pracy. Lu i współ. przedstawili pracę, w której omówiono zastosowanie modelu wieloetykietowego gradientowego drzewa decyzyjnego do przewidywania ryzyka zgonu z dowolnej przyczyny, udaru niedokrwienego oraz poważnych krwawień u pacjentów z AF ⁸². Do uczenia modelu posłużyły dane z rejestru GLORIA-AF ⁸³. Autorzy osiągnęli zadowalające wyniki AUC dla przewidywania zgonów z dowolnej przyczyny na poziomie 0.785 i porównali to z wynikami dla CCI wynoszącymi 0.747, model charakteryzował się o 10% lepszym wskaźnikiem poprawy netto. Wydajność modelu była spójna w różnych podgrupach, niezależnie od wieku, statusu leczenia przeciwzakrzepowego czy regionu geograficznego. Analiza istotności cech wykazała, że najważniejszymi predyktorami były: wiek, palenie tytoniu, klirens kreatyniny, stosowanie terapii

przeciwwzakrzepowej oraz przebyte udary lub przemijające ataki niedokrwienne.

Tabela 9. Porównanie wyników uzyskanych przez model z Publikacji II z innymi dostępnymi w literaturze

Parametr	Liczebność grupy	Przewidywanie	Algorytm	AUC	XAI
Publikacja II ⁸⁴	6431	Zgon w 180 dni w próbie pacjentów stosujących DOAC	LightGBM	0.886	SHAP
Chen i współ. ⁷⁷	1775	Przeżycie w 180 dni	LASSO z regresją Cox'a	0.788	Nomogram
Lv i współ. ⁷⁸	18727	Zgon z przyczyn sercowych w trakcie pobytu w szpitalu	XGBoost	0.932	SHAP
Nguyen i współ. ⁷⁹	20058 (trening); 11475 (test)	Zgon wśród pacjentów hospitalizowanych na oddziale intensywnej opieki medycznej	XGBoost	0.743	SHAP
Bernardini i współ. ⁸¹	11078 (wszyscy); 5943 (DOAC)	Zgon ze wszystkich przyczyn i zgon z przyczyn	Multi-Gate Mixture of Experts	0.733	SHAP – dla modelu GBDT

		sercowo-naczyniowych w próbie pacjentów stosujących DOAC			
Lu i współ. 82	25656	Zgon z dowolnej przyczyny	Wieloetykietykowane gradientowe drzewo decyzyjne	0.785	Istotność cech

W Tabeli 9 przedstawiono porównanie modeli predykcyjnych dla pacjentów z migotaniem przedsionków z wspomnianych prac z prezentowanym w niniejszej pracy rozwiązaniem. Jedynie model zaproponowany przez Lv i współ.⁷⁸ osiągnął wyższą wartość AUC niż model opracowany w niniejszej pracy. Warto jednak zaznaczyć, że model Lv i współ. obejmował relatywnie krótkie okno predykcji, ograniczone jedynie do okresu hospitalizacji, co może sprzyjać uzyskaniu wyższej skuteczności predykcyjnej ze względu na uzględnianie odmiennych czynników ryzyka. Prezentowany model natomiast przewiduje ryzyko zgonu w 180-dniowym okresie po wdrożeniu terapii DOAC, co stanowi większe wyzwanie analityczne. Wśród analizowanych publikacji jedynie model zaproponowany przez Chen i współ.⁷⁷ przewidywał przeżycie w 180-dniowym okresie obserwacji, co czyni go najbardziej porównywalnym z modelem opracowanym w niniejszej pracy pod względem horyzontu czasowego. Model Chen i współ. osiągnął wartość AUC równą 0,788, co jest niższym wynikiem w porównaniu uzyskanego przez prezentowany model oparty na algorytmie LightGBM. Należy również zwrócić uwagę, że model Chen i współ. wykorzystywał inne podejście analityczne (LASSO z regresją Coxa). Model zaproponowany przez Bernardini i współ.⁸¹ odnosił się również do populacji pacjentów stosujących DOAC,

ale charakteryzował się słabszymi wynikami. Większość pozostałych analizowanych publikacji obejmowała inne definicje zmiennych docelowych, co utrudnia bezpośrednie porównania

W ostatnich latach opublikowano niewiele prac poświęconych predykcji śmiertelności u pacjentów z tętniczym nadciśnieniem płucnym wykorzystując nowoczesne metody sztucznej inteligencji, zdecydowana większość publikacji dotyczy diagnostyki. Benza i współ. przeanalizowali podgrupę pacjentów z rejestru REVEAL⁸⁵, wykazując, że model REVEAL 2.0 osiąga AUC równe 0.76, co pozwala na skuteczną klasyfikację pacjentów do grup ryzyka i przewidywanie zarówno pogorszenia klinicznego, jak i śmiertelności w rocznej obserwacji⁸⁶. Model ten przewyższał pod względem zdolności dyskryminacyjnej strategię COMPERA i FPHR osiągające wyniki odpowiednio 0.62 i 0.64. Kramer i współ. w swoim przeglądzie podkreślili, że choć wiele modeli prognostycznych opiera się na danych obrazowych i klinicznych, rzadko są one walidowane zewnątrznie lub testowane prospektywnie³². Dodatkowo wykazano, że jedynie 9 z 53 uwzględnionych w przeglądzie prac, skupiało się na przewidywaniu śmiertelności lub innych punktów końcowych. Praca Alabed i współ. wykazała, że zastosowanie wieloliniowej analizy głównych składowych (MPCA) w obrazach MRI pozwala na poprawę predykcji rocznej śmiertelności (c-index 0.76 vs. 0.71 dla REVEAL)⁸⁷. W badaniu wykorzystano dane pochodzące z rejestru ASPIRE⁸⁸ do uczenia modelu opartego o algorytm maszyn wektorów nośnych (SVM). Największe ryzyko śmiertelności wiązało się z nieprawidłowościami w przegrodzie międzykomorowej w końcu skurczu oraz w lewej komorze w końcu rozkurczu. Autorzy podkreślili, że metoda ta pozwala na transparentną i klinicznie interpretowalną wizualizację cech

prognostycznych serca. Kheifets i współ. opracowali model lasu losowego do przewidywania 4-letniego ryzyka zgonu lub przeszczepu w kohorcie 167 dorosłych pacjentów z PAH, który osiągnął AUC 0.94 w walidacji wewnętrznej i 0.81 w zewnętrznej ⁸⁹. W uczeniu modelu wykorzystano zbiór danych z pracy Sweatt i współ. dotyczącej identyfikacji fenotypów immunologicznych w PAH stosując analizę proteomiczną ⁹⁰. Autorzy wykorzystali rekursywną eliminację cech do zminimalizowania wymiarowości danych. Analiza wartości SHAP pozwoliła na identyfikację najważniejszych predyktorów m.in. wynik testu 6-minutowego marszu, poziom NT-proBNP oraz biomarkery zapalne, takie jak interleukina 2 i 9. Autorzy podkreślili potrzebę rozwoju indywidualnych strategii fenotypowania oraz integracji danych klinicznych i biochemicznych z wynikami leczenia. Kanwar i współ. przedstawili drzewiasto rozszerzony naiwny klasyfikator Bayesa (tree-augmented naïve (TAN) Bayes) PHORA, w oparciu o rejestr REVEAL do przewidywania rocznego przeżycia u pacjentów z PAH, który uzyskał AUC 0.80 i wykazał przewagę nad REVEAL 2.0 (AUC 0.76). W walidacji zewnętrznej na danych z rejestru COMPERA ⁹¹ i PHSANZ ⁹² uzyskał AUC odpowiednio 0.74 i 0.80. Model ten wykazał bardzo dobrą separację grup ryzyka i lepszą zdolność dyskryminacyjną niż dotychczasowe narzędzia, co odzwierciedla możliwość modeli bayesowskich do uwzględniania wzajemnych zależności między zmiennymi klinicznymi oraz tolerancję na brakujące dane. Z kolei Suvon i współ. zaproponowali złożone podejście multimodalne, integrując cechy obrazowe, kategoryczne i tekstowe z elektronicznej dokumentacji medycznej, co pozwoliło osiągnąć wysoką skuteczność predykcji zgonu w ciągu roku (AUC 0.89) ⁹³. W pracy wykorzystano rejestr ASPIRE z którego włączone dane dla 2563

pacjentów. Wyniki te potwierdzają, że połączenie różnych źródeł danych znacząco poprawia trafność predykcji w PAH. Warto zwrócić uwagę, że autorzy zmagali się z dużym procentem brakujących danych i cech z ponad 40% brakiem zostały usunięte, a do pozostałych zastosowano nieliniową iteracyjną metodę najmniejszych kwadratów cząstkowych do imputacji danych. Użycie w pracy podejścia multimodalnego i wielowarstwowych perceptronów ogranicza znacząco możliwości łatwej interpretacji uzyskanych wyników oraz ich wytłumaczalności. Przegląd Sogol i współ. podkreśla, że AI ma ogromny potencjał w rewolucjonizowaniu diagnostyki, leczenia i prognozowania nadciśnienia płucnego, zwłaszcza w kontekście trudności wczesnego rozpoznania tej choroby ⁹⁴. Wyniki tych badań wskazują, że skuteczność modeli predykcyjnych w PAH zależy od jakości i różnorodności danych wejściowych oraz zastosowanych metod analitycznych, a integracja wielu źródeł informacji i nowoczesnych technik AI może znacząco poprawić trafność predykcji i wspierać personalizację opieki nad pacjentami z PAH.

Tabela 10. Porównanie wyników uzyskanych przez model z Publikacji III z innymi dostępnymi w literaturze

Parametr	Liczebność grupy	Przewidywanie	Algorytm	AUC	XAI
Publikacja III ⁹⁵	1755	Zgon do następnej wizyty z przyczyn sercowo-naczyniowych	XGBoost	0.767	SHAP

Benza i współ. ⁸⁶	2529	Roczna śmiertelność	Wieloczynnikowy model proporcjonalnego hazardu Cox'a	0.76	n/d - Kalkulator ryzyka
Kanwar i współ. ⁵⁹	2529	Roczne przeżycie	Model bayesowski	0.80/ 0.74	n/d - Kalkulator ryzyka
Alabed i współ. ⁸⁷	723	Roczna śmiertelność	Wieloliniowa analiza głównych składowych + maszyna wektorów nośnych	0.76	brak
Kheyfets i współ. ⁸⁹	167	4-letnie ryzyko zgonu	Las losowy	0.94/ 0.81	SHAP
Suvon i współ. ⁹³	2563	Roczna śmiertelność	Model multimodalny z wielowarstwowym perceptronem	0.89	brak

W Tabeli 10 przedstawiono porównanie modeli predykcyjnych dla pacjentów z tętniczym nadciśnieniem płucnym z wspomnianych prac z prezentowanym w niniejszej pracy rozwiązaniem. Wyniki uzyskane w niniejszej pracy (AUC 0,767 dla modelu XGBoost przewidującego zgon z przyczyn sercowo-naczyniowych do następnej wizyty) są porównywalne z rezultatami innych publikacji dotyczących predykcji ryzyka w populacji pacjentów z tętniczym nadciśnieniem płucnym. Modele Benza i współ. ⁸⁶ (AUC 0,76) oraz Alabed i współ. ⁸⁷ (AUC 0,76) osiągnęły bardzo zbliżoną skuteczność predykcyjną, mimo zastosowania innych algorytmów i danych wejściowych. Model Kanwara i współ. ⁵⁹ wykazał nieco wyższą wartość AUC (0,80 dla rocznego przeżycia), natomiast model Suvon i współautorów, wykorzystujący multimodalny perceptron, osiągnął

najwyższą skuteczność (AUC 0,89), jednak stracił sporo na wytłumaczalności przez zastosowanie sieci neuronowej. Z kolei model Kheifets i współ.⁸⁹ uzyskał bardzo wysoką wartość AUC (0,94), lecz dotyczył czteroletniego ryzyka zgonu w znacznie mniejszej kohorcie. Warto podkreślić, że dla zbioru zewnętrznej walidacji wynik AUC był niższy i wynosił 0.81, co przyrównuje go do pozostałych przykładów. Autorzy jako jedyni z przytoczonych prac wykorzystali również wartości SHAP dla wytłumaczalności uzyskiwanych rezultatów. Wyniki te wskazują, że skuteczność predykcyjna opracowanego modelu jest porównywalna z uznanymi narzędziami prognostycznymi, a zastosowanie algorytmu XGBoost oraz metody SHAP zapewnia zarówno wysoką jakość predykcji, jak i interpretowalność wyników.

Należy jednak podkreślić, że pomimo rosnącej skuteczności modeli predykcyjnych, ich wdrożenie w praktyce klinicznej wymaga dalszej walidacji zewnętrznej, integracji danych z różnych źródeł oraz ścisłej współpracy interdyscyplinarnej. Jak wskazują autorzy cytowanych badań, kluczowe jest również zapewnienie wysokiej jakości danych wejściowych oraz precyzyjnego etykietowania, co znacząco wpływa na moc prognostyczną modeli. Modele oparte na pojedynczej modalności, choć wciąż użyteczne, ustępują miejsca rozwiązaniom integrującym wiele źródeł informacji, co pozwala na lepsze uchwycenie złożoności klinicznej PAH. Warto również zwrócić uwagę na rosnącą rolę interpretowalności modeli AI w medycynie. Najnowsze prace pokazują, że nowoczesne podejście umożliwia nie tylko uzyskanie wysokiej skuteczności predykcyjnej, ale także identyfikację kluczowych czynników ryzyka, co ma istotne znaczenie dla praktyki klinicznej. W kontekście przedstawionego modelu, osiągającego AUC 0.767, można stwierdzić, że

wpisuje się on w aktualny trend rozwoju narzędzi opartych na sztucznej inteligencji, które mogą stanowić cenne wsparcie w codziennej praktyce klinicznej, zwłaszcza w tak trudnej i złożonej jednostce chorobowej jak PAH.

Tabela 11. Zestawienie parametrów modeli z publikacji II i III

Parametr	Publikacja II	Publikacja III
Jednostka chorobowa	AF	PAH
Liczebność grupy	6431	1755
Balans klas	1:17.64	1:12.93
Przewidywany zgon w czasie	6 miesięcy	średnio 12 miesięcy
Algorytm	LightGBM	XGBoost
Ilość cech do budowy modelu	27	17
Dokładność	0.862 (0.842–0.880)	0.738 (0.695–0.783)
Czułość	0.913 (0.842–0.973)	0.800 (0.636–0.947)
Specyficzność	0.859 (0.839–0.878)	0.733 (0.688–0.779)
MCC	0.450 (0.388–0.507)	0.298 (0.195–0.399)
AUC	0.886 (0.849–0.919)	0.767 (0.682–0.843)
XAI	SHAP	SHAP

W Tabela 11 przedstawiono porównanie kluczowych parametrów oraz wyników predykcyjnych modeli uczenia maszynowego opracowanych dla dwóch odmiennych jednostek chorobowych: migotania przedsionków oraz tętniczego nadciśnienia płucnego. Modele różniły się zarówno liczebnością analizowanych grup, jak i liczbą wykorzystanych cech oraz okresem obserwacji. W obu przypadkach zastosowano

zaawansowane algorytmy wzmocnienia gradientowego (LightGBM dla AF oraz XGBoost dla PAH) oraz metodę SHAP do interpretacji wyników.

Wyniki wskazują na wyższą skuteczność predykcyjną modelu opracowanego dla pacjentów z AF, co potwierdzają wyższe wartości wszystkich metryk wydajności, dokładności (0,862 vs. 0,738), czułości (0,913 vs. 0,800), specyficzności (0,859 vs. 0,733), MCC (0,450 vs. 0,298) oraz AUC (0,886 vs. 0,767), w porównaniu do modelu dla PAH. Różnice w skuteczności predykcyjnej modeli mogą wynikać z wielu czynników. Przede wszystkim, populacje pacjentów z migotaniem przedsionków i tętniczym nadciśnieniem płucnym różnią się pod względem charakterystyki klinicznej, etiologii, patofizjologii, częstości występowania chorób współistniejących oraz naturalnego przebiegu choroby, co może wpływać na trudność predykcji. Ponadto, liczba i jakość dostępnych danych, a także różnice w czasie obserwacji i definicjach punktów końcowych, mogły mieć istotny wpływ na wydajność modeli. Przewidywanie w dłuższym okresie czasu oraz zmienność okresu obserwacji w danych uczących dla modelu PAH znacząco wpływa na uzyskiwane wyniki. Istotne znaczenie miały również wielkość i balans klas w analizowanych zbiorach danych. Mała próba i duże niebalansowanie klas mogą znacząco ograniczać jakość wyników. Wszystkie te czynniki mogą przyczyniać się do obserwowanych rozbieżności w wynikach modeli, dlatego interpretując uzyskane rezultaty należy zachować ostrożność i uwzględnić specyfikę każdej z badanych populacji. Należy jednak podkreślić, że bezpośrednie porównywanie wyników pomiędzy modelami dla różnych jednostek chorobowych jest ograniczone.

Oba przedstawione modele wykazały się dobrą skutecznością predykcyjną w swoich kohortach, a zastosowanie metod wyjaśnialnej

sztucznej inteligencji umożliwiło identyfikację najważniejszych czynników ryzyka. Wyniki te potwierdzają przydatność zaawansowanych narzędzi analitycznych w personalizacji oceny ryzyka w kardiologii, przy jednoczesnym zachowaniu ostrożności w interpretacji porównań między różnymi jednostkami chorobowymi.

5.1. Podsumowanie

Podsumowując, AI ma potencjał, by zrewolucjonizować kardiologię, poprawiając trafność diagnoz, umożliwiając personalizację leczenia i wspierając podejmowanie decyzji klinicznych. Jednak jej wdrożenie wymaga ścisłej współpracy interdyscyplinarnej, jasnych wytycznych etycznych oraz ciągłej oceny jakości i bezpieczeństwa danych. Ostatecznie, narzędzia AI powinny być traktowane jako wsparcie dla lekarzy, a nie jako autonomiczne systemy decyzyjne, a ich skuteczność i bezpieczeństwo muszą być potwierdzone w szeroko zakrojonych badaniach klinicznych.

5.2. Ograniczenia

W przedstawionych pracach, istotnym ograniczeniem jest dostępność i jakość danych – np. w bazie MIMIC-IV liczba pacjentów leczonych DOAC w AF jest niska ze względu na okres gromadzenia danych, a dane pochodzą z jednego ośrodka, co ogranicza możliwość generalizacji wyników. W drugim badaniu ograniczeniem jest niewielka liczebność próby pacjentów z PAH oraz jej jednorodność populacyjna, co może prowadzić do błędów predykcyjnych na zewnętrznych zbiorach danych.

Dodatkowo, uproszczone podejście do braków danych (usuwanie niekompletnych rekordów) może wprowadzać bias, a zastosowanie bardziej zaawansowanych metod imputacji nie przynosiło poprawy

wyników. Modele predykcyjne nie rozróżniają przyczyn zgonu, a wybrane cechy (np. stosowanie leków) odzwierciedlają raczej stan kliniczny pacjenta niż bezpośredni wpływ na ryzyko śmiertelności przez co wymagają doświadczenia w poprawnej interpretacji.

Ważnym aspektem jest także ryzyko fałszywie dodatnich i fałszywie ujemnych predykcji – nadmierna liczba fałszywie dodatnich może prowadzić do niepotrzebnych interwencji i obciążenia systemu ochrony zdrowia, natomiast fałszywie ujemne mogą skutkować przeoczeniem pacjentów wymagających intensywniejszej opieki. Z tego względu wyniki modeli AI powinny być traktowane jako narzędzie wspierające decyzje kliniczne, a nie zastępujące doświadczenie lekarza.

5.3. Ograniczenia metodologiczne

Uczenie maszynowe znajduje coraz szersze zastosowanie w medycynie, jednak jego rozwój i praktyczna użyteczność są nadal istotnie ograniczone przez szereg problemów o charakterze algorytmicznym i metodologicznym. Na dzień dzisiejszy jednym z kluczowych wyzwań pozostaje jakość oraz dostępność danych medycznych, które często są ograniczone pod względem ilościowym, dodatkowo są niejednorodne oraz obarczone błędami anotacji wynikającymi z subiektywności ekspertów klinicznych. Dane medyczne pochodzące z elektronicznej dokumentacji medycznej rzadko są ustrukturyzowane co zwiększa nakład pracy oraz wymaga stosowania dodatkowych narzędzi. W przeciwieństwie do wielu innych dziedzin, dane medyczne są trudne i kosztowne do pozyskania, a dodatkowo podlegają restrykcjom prawnym, co ogranicza możliwość budowy dużych, reprezentatywnych zbiorów treningowych. W konsekwencji modele uczące się są szczególnie podatne na przeuczenie (overfitting) oraz problemy z generalizacją, zwłaszcza w sytuacjach

zmiany rozkładu danych (domain shift), np. pomiędzy różnymi ośrodkami klinicznymi, populacjami pacjentów czy urządzeniami diagnostycznymi. Kolejnym istotnym ograniczeniem jest brak interpretowalności złożonych modeli, w szczególności głębokich sieci neuronowych, które funkcjonują jako tzw. „czarne skrzynki”. W kontekście medycznym, gdzie decyzje diagnostyczne i terapeutyczne wymagają uzasadnienia oraz podlegają rygorystycznym regulacjom, brak przejrzystości modelu znacząco utrudnia jego wdrożenie i akceptację przez personel medyczny. Zaufanie do rozwiązań opartych o AI jest fundamentalne dla zrewolucjonizowania systemu ochrony zdrowia, a brak możliwości przedstawienia wyjaśnialności dla prezentowanego wyniku rodzi uzasadnione obawy całego środowiska medycznego. Równocześnie większość współczesnych metod uczenia maszynowego opiera się na identyfikacji zależności korelacyjnych, nie zaś przyczynowych, co ogranicza ich zdolność do wspierania decyzji klinicznych wymagających zrozumienia mechanizmów biologicznych oraz przewidywania efektów interwencji. Problem ten jest szczególnie widoczny w zadaniach związanych z planowaniem leczenia, gdzie konieczne jest rozróżnienie między korelacją, a rzeczywistym wpływem czynników na stan zdrowia pacjenta. Z tego powodu interpretacja wyników modeli AI/ML wymaga zrozumienia na jakie zasadzie powstają przewidywania i rekomendacje modelu oraz co może na nie wpłynąć. Istotnym wyzwaniem pozostaje również właściwe modelowanie niepewności predykcji. W rozwiązaniach opartych o AI/ML wiemy, że dojdzie do błędnych prognoz, jednak musimy wyznaczać granice jakie błędy jesteśmy w stanie zaakceptować oraz kiedy ryzyko jest zbyt duże. Wiele algorytmów dostarcza deterministycznych wyników bez wiarygodnej estymacji niepewności, co jest nieadekwatne w środowisku

klinicznym, gdzie kluczowe znaczenie ma identyfikacja przypadków wymagających dodatkowej weryfikacji. Dodatkowo dane medyczne mają charakter wielomodalny, obejmując obrazy diagnostyczne, dane tekstowe, sygnały czasowe oraz informacje genetyczne, co stawia przed algorytmami wyzwania związane z efektywną integracją heterogenicznych źródeł informacji. Pomimo postępów w projektowaniu architektur wielomodalnych, ich skuteczność nadal silnie zależy od dostępności dużych i dobrze zsynchronizowanych zbiorów danych. Kolejnym problemem jest ograniczona stabilność i reprodukowalność modeli – niewielkie zmiany w danych treningowych lub procedurze uczenia mogą prowadzić do istotnych różnic w wynikach, co utrudnia walidację i porównywanie metod. Wreszcie, stosowane metody optymalizacyjne koncentrują się na maksymalizacji metryk wydajności, takich jak dokładność czy AUC, które nie zawsze przekładają się na rzeczywiste korzyści kliniczne. Brak odpowiednio zaprojektowanych badań prospektywnych oraz walidacji w warunkach rzeczywistej praktyki klinicznej dodatkowo ogranicza wiarygodność uzyskiwanych rezultatów. W efekcie, mimo dynamicznego rozwoju metod uczenia maszynowego, ich zastosowanie w medycynie pozostaje obszarem wymagającym dalszych badań nad bardziej odpornymi, interpretowalnymi i przyczynowo ugruntowanymi modelami, zdolnymi do działania w warunkach wysokiej niepewności i złożoności danych.

Istotnym, często niedostatecznie eksponowanym ograniczeniem współczesnych metod uczenia maszynowego w medycynie są również uwarunkowania obliczeniowe oraz specyfika zastosowania dużych modeli językowych, które w ostatnich latach stały się jednym z dominujących paradygmatów w obszarze AI. W przypadku zaawansowanych modeli, w

szczegółności tych należących do nurtu uczenia głębokiego, wymagania obliczeniowe związane z procesem trenowania i wnioskowania są bardzo wysokie, co ogranicza ich dostępność oraz możliwość zastosowania w środowiskach klinicznych. Trenowanie modeli o dużej liczbie parametrów wymaga dostępu do wyspecjalizowanej kosztownej infrastruktury, znacznych zasobów energetycznych oraz czasu, co stanowi barierę zarówno ekonomiczną, jak i organizacyjną. W praktyce prowadzi to do koncentracji rozwoju najbardziej zaawansowanych modeli w niewielkiej liczbie ośrodków badawczych i firm technologicznych, co może utrudniać ich niezależną walidację oraz adaptację do specyficznych potrzeb lokalnych systemów ochrony zdrowia. Ponadto, wysokie koszty obliczeniowe utrudniają częste aktualizowanie modeli, co jest istotne w kontekście dynamicznie zmieniającej się wiedzy medycznej oraz zjawiska dryfu danych.

W kontekście zastosowań przetwarzania języka naturalnego w medycynie, szczególną uwagę zwracają duże modele językowe, które umożliwiają analizę i generowanie tekstu medycznego, wspierając m.in. przetwarzanie dokumentacji klinicznej, ekstrakcję informacji wraz z ich strukturyzacją czy systemy wspomagania decyzji. Niemniej jednak ich wykorzystanie wiąże się z szeregiem ograniczeń metodologicznych. LLM operują na poziomie statystycznych wzorców językowych, co oznacza, że mogą generować odpowiedzi pozornie poprawne, lecz niezgodne z aktualną wiedzą medyczną (tzw. halucynacje). Problem ten jest szczególnie istotny w zastosowaniach klinicznych, gdzie nawet drobne błędy mogą prowadzić do poważnych konsekwencji, co zaś tworzy dylematy natury prawnej oraz etycznej. Dodatkowo modele te mają ograniczoną zdolność do weryfikacji prawdziwości generowanych treści

oraz do rozumienia kontekstu klinicznego w sposób odpowiadający rzeczywistemu rozumowaniu lekarza. Kolejnym wyzwaniem jest ich wrażliwość na sposób formułowania zapytań, co może prowadzić do niespójnych wyników dla semantycznie podobnych wejść. Z perspektywy obliczeniowej, wnioskowanie z wykorzystaniem dużych modeli językowych również może być kosztowne i czasochłonne, co ogranicza ich zastosowanie w systemach wymagających działania w czasie rzeczywistym. Dodatkowo, ze względu na rozmiar modeli i charakter danych treningowych, ich adaptacja do wyspecjalizowanych dziedzin medycznych wymaga stosowania technik takich jak fine-tuning lub retrieval-augmented generation, które z kolei wprowadzają dodatkową złożoność systemową oraz wymagania infrastrukturalne, co dla małych ośrodków badawczych tworzy dodatkowe bariery.

Nie bez znaczenia pozostają również kwestie związane z prywatnością i bezpieczeństwem danych. Wdrażanie modeli, zwłaszcza w architekturze chmurowej, może kolidować z regulacjami dotyczącymi ochrony danych medycznych, co ogranicza możliwość wykorzystania pełnego potencjału obliczeniowego dostępnego poza lokalną infrastrukturą placówek medycznych. Alternatywne podejścia, takie jak federacyjne uczenie modeli, redukują konieczność centralizacji danych, jednak wprowadzają dodatkowe wyzwania związane z synchronizacją, stabilnością procesu uczenia oraz efektywnością komunikacyjną. W rezultacie ograniczenia obliczeniowe oraz specyfika dużych modeli językowych stanowią istotny czynnik hamujący ich szerokie i bezpieczne wdrożenie w praktyce klinicznej, podkreślając potrzebę dalszych badań nad bardziej efektywnymi, wiarygodnymi i energooszczędnymi metodami przetwarzania danych medycznych.

5.4. Szanse aplikacyjne

Przedstawione w pracy rozwiązanie ma istotne szanse aplikacyjne. Wymaga jednak opracowania intuicyjnego narzędzia przeznaczonego dla użytkownika końcowego, którym jest lekarz. W dłuższej perspektywie wdrożenie AI może przyczynić się do zwiększenia efektywności systemu ochrony zdrowia, ograniczenia kosztów oraz wyrównania dostępu do nowoczesnych metod diagnostycznych i terapeutycznych, zwłaszcza w regionach o ograniczonych zasobach kadrowych i infrastrukturalnych. Integracja takich rozwiązań z codzienną praktyką kliniczną może również poprawić jakość opieki nad pacjentem poprzez wsparcie procesu decyzyjnego i umożliwienie bardziej spersonalizowanego podejścia terapeutycznego.

6. Wnioski

- 1) Modele uczenia maszynowego, takie jak LightGBM i XGBoost, wykazują wysoką skuteczność w przewidywaniu śmiertelności pacjentów z chorobami sercowo-naczyniowymi, takimi jak AF i PAH.
- 2) W obu badaniach zastosowano metodę SHAP do wyjaśnienia predykcji modeli i oceny znaczenia poszczególnych cech. Interpretowalność wyników pozwala na identyfikację kluczowych czynników ryzyka dla indywidualnych pacjentów, co zwiększa praktyczną użyteczność modeli w klinice oraz pozwala lepiej zrozumieć na jakich cechach opiera się predykcja co zwiększa zaufanie do uzyskanych wyników.
- 3) Modele osiągnęły wysokie wartości metryk, takich jak AUC, czułość i swoistość, co potwierdza ich skuteczność w przewidywaniu śmiertelności.
- 4) Modele te mogą wspierać tradycyjne metody oceny ryzyka, oferując bardziej szczegółowe i spersonalizowane podejście do zarządzania pacjentami. Zastosowanie algorytmów uczenia maszynowego może przyczynić się do poprawy wyników leczenia i zmniejszenia obciążenia systemu opieki zdrowotnej w kontekście ekonomicznym oraz obciążenia pracą.
- 5) Wyniki badań wskazują na potencjał dalszego rozwoju modeli predykcyjnych, w tym ich walidacji na innych populacjach i w różnych warunkach klinicznych.

7. Piśmiennictwo

1. Reinsel D, Gantz J, Rydning J. Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data; Focus on the Data That's Big. *Framingham: International Data Corporation*. Published online April 2017:1-25. www.idc.com
2. Reinsel D, Gantz J, Rydning J. The Digitization of the World From Edge to Core. *Framingham: International Data Corporation*. 2020;(16):1-28.
3. Bahri S, Zoghلامي N, Abed M, Tavares JMRS. BIG DATA for Healthcare: A Survey. *IEEE Access*. 2019;7:7397-7408. doi:10.1109/ACCESS.2018.2889180
4. McCarthy J, MML, RN, & SCE. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Mag*. 1955;27(4):12. doi:<https://doi.org/10.1609/aimag.v27i4.1904>
5. TURING AM. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*. 1950;LIX(236):433-460. doi:10.1093/mind/LIX.236.433
6. Stark BA, DeCleene NK, Desai EC, et al. Global, Regional, and National Burden of Cardiovascular Diseases and Risk Factors in 204 Countries and Territories, 1990-2023. *JACC*. 2025;86(22):2167-2243. doi:10.1016/j.jacc.2025.08.015
7. Celestin B, Bagherzadeh SP, Santana E, et al. Artificial Intelligence-Based Echocardiography in Pulmonary Arterial Hypertension. *Chest*. 2026;169(1):207-219. doi:10.1016/j.chest.2025.06.052
8. Salehi M, Alabed S, Sharkey M, et al. Artificial intelligence-based echocardiography assessment to detect pulmonary hypertension. *ERJ Open Res*. 2025;11(3):00592-02024. doi:10.1183/23120541.00592-2024

9. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25(1):70-74. doi:10.1038/s41591-018-0240-2
10. Croon PM, Selder JL, Allaart CP, et al. Current state of artificial intelligence-based algorithms for hospital admission prediction in patients with heart failure: A scoping review. *European Heart Journal - Digital Health*. Oxford University Press. 2022;3(3):415-425. doi:10.1093/ehjdh/ztac035
11. Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial Intelligence in Cardiology. *J Am Coll Cardiol*. 2018;71(23):2668-2679. doi:10.1016/J.JACC.2018.03.521
12. Watson X, D'Souza J, Cooper D, Markham R. Artificial intelligence in cardiology: fundamentals and applications. *Intern Med J*. 2022;52(6):912-920. doi:10.1111/IMJ.15562
13. Westcott RJ, Tcheng JE. Artificial Intelligence and Machine Learning in Cardiology. *JACC Cardiovasc Interv*. 2019;12(14):1312-1314. doi:10.1016/J.JCIN.2019.03.026
14. Krittanawong C, Zhang HJ, Wang Z, Aydar M, Kitai T. Artificial Intelligence in Precision Cardiovascular Medicine. *J Am Coll Cardiol*. 2017;69(21):2657-2664. doi:10.1016/J.JACC.2017.03.571
15. Institute for Health Metrics and Evaluation. Global Burden of Disease Study 2023. <https://ghdx.healthdata.org/>. 2023. Accessed March 24, 2026. <https://ghdx.healthdata.org/>
16. Van Gelder IC, Rienstra M, Bunting K V, et al. 2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J*. 2024;45(36):3314-3414. doi:10.1093/eurheartj/ehae176

17. Hu WS, Lin CL. Comparisons of HATCH, HAVOC and CHA2DS2-VASc scores for all-cause mortality prediction in atrial fibrillation: a real-world evidence study. *Postgrad Med J*. 2023;99(1170):326-332. doi:10.1136/postgradmedj-2021-141147
18. Hu WS, Li Lin C. Real-world observational study of assessment of CHA2DS2-VASc, C2HEST and HAVOC scores for atrial fibrillation among patients with rheumatological disorders: a nationwide analysis. *Postgrad Med J*. 2022;98(1165):837-841. doi:10.1136/postgradmedj-2021-140754
19. Proietti M, Farcomeni A, Romiti GF, et al. Association between clinical risk scores and mortality in atrial fibrillation: Systematic review and network meta-regression of 669,000 patients. *Eur J Prev Cardiol*. 2020;27(6):633-644. doi:10.1177/2047487318817662
20. Fox KAA, Virdone S, Pieper KS, et al. GARFIELD-AF risk score for mortality, stroke, and bleeding within 2 years in patients with atrial fibrillation. *Eur Heart J Qual Care Clin Outcomes*. 2022;8(2):214-227. doi:10.1093/ehjqcco/qcab028
21. Wang B, Jin F, Cao H, Li Q, Zhang P. Development of a risk score for predicting one-year mortality in patients with atrial fibrillation using XGBoost-assisted feature selection. *Polish Heart Journal*. 2024;82(10):941-948. doi:10.33963/v.phj.101842
22. Liu X, Jiang J, Wei L, et al. Prediction of all-cause mortality in coronary artery disease patients with atrial fibrillation based on machine learning models. *BMC Cardiovasc Disord*. 2021;21(1):499. doi:10.1186/s12872-021-02314-w
23. Falsetti L, Rucco M, Proietti M, et al. Risk prediction of clinical adverse outcomes with machine learning in a cohort of critically ill patients with atrial fibrillation. *Sci Rep*. 2021;11(1):18925. doi:10.1038/s41598-021-97218-2

24. Humbert M, Kovacs G, Hoeper MM, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J*. 2022;43(38):3618-3731. doi:10.1093/eurheartj/ehac237
25. Kaemmerer H, Diller GP, Dähnert I, et al. Pulmonary hypertension in adults with congenital heart defects (ACHDs) in light of the 2022 ESC PAH guidelines—part II: supportive therapy, special situations (pregnancy, contraception, non-cardiac surgery), targeted pharmacotherapy, organ transplantation, special management (shunt lesion, left ventricular disease, univentricular hearts), interventions, intensive care, ACHD follow-up, future perspective. *Cardiovasc Diagn Ther*. 2024;14(5):921-934. doi:10.21037/cdt-24-167
26. Benza RL, Miller DP, Gomberg-Maitland M, et al. Predicting Survival in Pulmonary Arterial Hypertension. *Circulation*. 2010;122(2):164-172. doi:10.1161/CIRCULATIONAHA.109.898122
27. Benza RL, Gomberg-Maitland M, Miller DP, et al. The REVEAL Registry Risk Score Calculator in Patients Newly Diagnosed With Pulmonary Arterial Hypertension. *Chest*. 2012;141(2):354-362. doi:10.1378/chest.11-0676
28. Hoeper MM, Kramer T, Pan Z, et al. Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. *European Respiratory Journal*. 2017;50(2):1700740. doi:10.1183/13993003.00740-2017
29. Kylhammar D, Kjellström B, Hjalmarsson C, et al. A comprehensive risk stratification at early follow-up determines prognosis in pulmonary arterial hypertension. *Eur Heart J*. 2018;39(47):4175-4181. doi:10.1093/eurheartj/ehx257
30. Boucly A, Weatherald J, Savale L, et al. Risk assessment, prognosis and guideline implementation in pulmonary arterial

- hypertension. *European Respiratory Journal*. 2017;50(2):1700889. doi:10.1183/13993003.00889-2017
31. Kraisangka J, Druzdzal MJ, Lohmueller LC, Kanwar MK, Antaki JF, Benza RL. Bayesian Network vs. Cox's Proportional Hazard Model of PAH Risk: A Comparison. In: 2019:139-149. doi:10.1007/978-3-030-21642-9_19
 32. Kramer T, Kramer M, Hagist C, Spinler S. Artificial intelligence in pulmonary hypertension: a systematic review. *Eur J Med Res*. 2025;30(1):1225. doi:10.1186/s40001-025-03557-5
 33. World Health Organization. *ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH*. 2021. Accessed October 17, 2022. <https://www.who.int/publications/i/item/9789240029200>
 34. Sengupta PP, Shrestha S, Berthon B, et al. Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): A Checklist. *JACC Cardiovasc Imaging*. 2020;13(9):2017-2035. doi:10.1016/j.jcmg.2020.07.015
 35. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *British Journal of Surgery*. 2015;102(3):148-158. doi:10.1002/bjs.9736
 36. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet*. 2020;395(10236):1579. doi:10.1016/S0140-6736(20)30226-9
 37. General Data Protection Regulation (GDPR) Compliance Guidelines. Accessed October 15, 2022. <https://gdpr.eu/>
 38. Morley J, Taddeo M, Floridi L. Google Health and the NHS: overcoming the trust deficit. *Lancet Digit Health*. 2019;1(8):e389. doi:10.1016/S2589-7500(19)30193-1

39. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. Guyon I, Luxburg U Von, Bengio S, et al., eds. *Adv Neural Inf Process Syst*. 2017;30. <http://arxiv.org/abs/1705.07874>
40. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67. doi:10.1038/s42256-019-0138-9
41. Bergstra J, Yamins D, Cox D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In: Dasgupta S, McAllester D, eds. *Proceedings of the 30th International Conference on Machine Learning*. Vol 28. Proceedings of Machine Learning Research. PMLR; 2013:115-123. <https://proceedings.mlr.press/v28/bergstra13.html>
42. Alistair J, Lucas B, Tom P, et al. MIMIC-IV (version 3.1). *PhysioNet*. Preprint posted online 2024. doi:10.13026/kpb9-mt58
43. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10(1):1. doi:10.1038/s41597-022-01899-x
44. Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*. 2000;101(23). doi:10.1161/01.CIR.101.23.e215
45. Kopeć G, Kurzyna M, Mroczek E, et al. Database of Pulmonary Hypertension in the Polish Population (BNP-PL): design of the registry. *Kardiol Pol*. 2019;77(10):972-974. doi:10.33963/KP.14988
46. Kwiatkowska J, Zuk M, Migdal A, et al. Children and Adolescents with Pulmonary Arterial Hypertension: Baseline and Follow-Up Data from the Polish Registry of Pulmonary Hypertension (BNP-PL). *J Clin Med*. 2020;9(6):1717. doi:10.3390/jcm9061717

47. Kopeć G, Dzikowska-Diduch O, Mroczek E, et al. Characteristics and outcomes of patients with chronic thromboembolic pulmonary hypertension in the era of modern therapeutic approaches: data from the Polish multicenter registry (BNP-PL). *Ther Adv Chronic Dis.* 2021;12. doi:10.1177/20406223211002961
48. Braksator M, Kurzyna M, Kopeć G, et al. Hemodynamic, echocardiographic, and demographic profiles of patients with chronic thromboembolic pulmonary hypertension and atrial fibrillation: A multicenter cohort study. *The Journal of Heart and Lung Transplantation.* 2025;44(8):1239-1248. doi:10.1016/j.healun.2025.03.020
49. Kopeć G, Kurzyna M, Mroczek E, et al. Characterization of patients with pulmonary arterial hypertension: Data from the polish registry of pulmonary hypertension (BNP-PL). *J Clin Med.* 2020;9(1). doi:10.3390/jcm9010173
50. Kogan E, Didden EM, Lee E, et al. A machine learning approach to identifying patients with pulmonary hypertension using real-world electronic health records. *Int J Cardiol.* 2023;374:95-99. doi:10.1016/j.ijcard.2022.12.016
51. Mohammad MA, Olesen KKW, Koul S, et al. Development and validation of an artificial neural network algorithm to predict mortality and admission to hospital for heart failure after myocardial infarction: a nationwide population-based study. *Lancet Digit Health.* 2022;4(1):e37-e45. doi:10.1016/S2589-7500(21)00228-4
52. Moghaddasi H, Nourian S. Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos. *Comput Biol Med.* 2016;73:47-55. doi:10.1016/j.compbiomed.2016.03.026
53. Porumb M, Stranges S, Pescapè A, Pecchia L. Precision Medicine and Artificial Intelligence: A Pilot Study on Deep Learning for

- Hypoglycemic Events Detection based on ECG. *Sci Rep*. 2020;10(1):170. doi:10.1038/s41598-019-56927-5
54. Salte IM, Østvik A, Smistad E, et al. Artificial Intelligence for Automatic Measurement of Left Ventricular Strain in Echocardiography. *JACC Cardiovasc Imaging*. 2021;14(10):1918-1928. doi:10.1016/j.jcmg.2021.04.018
 55. Kusunose K, Abe T, Haga A, et al. A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images. *JACC Cardiovasc Imaging*. 2020;13(2):374-381. doi:10.1016/j.jcmg.2019.02.024
 56. Berikol GB, Yildiz O, Özcan İT. Diagnosis of Acute Coronary Syndrome with a Support Vector Machine. *J Med Syst*. 2016;40(4):84. doi:10.1007/s10916-016-0432-6
 57. Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J*. 2016;38(7):ehw188. doi:10.1093/eurheartj/ehw188
 58. Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, Naghavi M. Machine Learning Outperforms ACC/AHA CVD Risk Calculator in MESA. *J Am Heart Assoc*. 2018;7(22). doi:10.1161/JAHA.118.009476
 59. Kanwar MK, Gomberg-Maitland M, Hoeper M, et al. Risk stratification in pulmonary arterial hypertension using Bayesian analysis. *European Respiratory Journal*. 2020;56(2):2000008. doi:10.1183/13993003.000008-2020
 60. Galloway CD, Valys A V., Shreibati JB, et al. Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia From the Electrocardiogram. *JAMA Cardiol*. 2019;4(5):428. doi:10.1001/jamacardio.2019.0640

61. Luongo G, Azzolin L, Schuler S, et al. Machine learning enables noninvasive prediction of atrial fibrillation driver location and acute pulmonary vein ablation success using the 12-lead ECG. *Cardiovasc Digit Health J*. 2021;2(2):126-136. doi:10.1016/j.cvdhj.2021.03.002
62. Karwath A, Bunting K V, Gill SK, et al. Redefining β -blocker response in heart failure patients with sinus rhythm and atrial fibrillation: a machine learning cluster analysis. *The Lancet*. 2021;398(10309):1427-1435. doi:10.1016/S0140-6736(21)01638-X
63. Cikes M, Sanchez-Martinez S, Claggett B, et al. Machine Learning-Based Phenogrouping in Heart Failure to Identify Responders to Cardiac Resynchronization Therapy. *Eur J Heart Fail*. 2019;21(1):74-85. doi:10.1002/ejhf.1333
64. Levy AE, Biswas M, Weber R, et al. Applications of machine learning in decision analysis for dose management for dofetilide. Rasmusson RL, ed. *PLoS One*. 2019;14(12):e0227324. doi:10.1371/journal.pone.0227324
65. Garvin JH, Kim Y, Gobbel GT, et al. Automating Quality Measures for Heart Failure Using Natural Language Processing: A Descriptive Study in the Department of Veterans Affairs. *JMIR Med Inform*. 2018;6(1):e5. doi:10.2196/medinform.9150
66. Shah RU, Mukherjee R, Zhang Y, et al. Impact of Different Electronic Cohort Definitions to Identify Patients With Atrial Fibrillation From the Electronic Medical Record. *J Am Heart Assoc*. 2020;9(5). doi:10.1161/JAHA.119.014527
67. Kaspar M, Fette G, Güder G, et al. Underestimated prevalence of heart failure in hospital inpatients: a comparison of ICD codes and discharge letter information. *Clinical Research in Cardiology*. 2018;107(9):778-787. doi:10.1007/s00392-018-1245-z

68. Patel YR, Robbins JM, Kurgansky KE, et al. Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records. *BMC Cardiovasc Disord.* 2018;18(1):128. doi:10.1186/s12872-018-0866-5
69. Mahajan Satish M., Ghani Rayid. Combining Structured and Unstructured Data for Predicting Risk of Readmission for Heart Failure Patients. In: *Studies in Health Technology and Informatics.* Vol 264. Stud Health Technol Inform; 2019:238-242. doi:10.3233/SHTI190219
70. Galper BZ, Beery DE, Leighton G, Englander LL. Comparison of adverse event and device problem rates for transcatheter aortic valve replacement and Mitraclip procedures as reported by the Transcatheter Valve Therapy Registry and the Food and Drug Administration postmarket surveillance data. *Am Heart J.* 2018;198:64-74. doi:10.1016/j.ahj.2017.10.013
71. Afzal N, Sohn S, Abram S, et al. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. *J Vasc Surg.* 2017;65(6):1753-1761. doi:10.1016/j.jvs.2016.11.031
72. Ashburner JM, Chang Y, Wang X, et al. Natural Language Processing to Improve Prediction of Incident Atrial Fibrillation Using Electronic Health Records. *J Am Heart Assoc.* 2022;11(15):26014. doi:10.1161/JAHA.122.026014
73. Camm AJ, Amarencio P, Haas S, et al. XANTUS: a real-world, prospective, observational study of patients treated with rivaroxaban for stroke prevention in atrial fibrillation. *Eur Heart J.* 2016;37(14):1145-1153. doi:10.1093/eurheartj/ehv466
74. LAVALLE C, PIERUCCI N, MARIANI M V., et al. Italian Registry in the Setting of Atrial Fibrillation Ablation with Rivaroxaban - IRIS. *Minerva Cardiology and Angiology.* 2024;72(6). doi:10.23736/S2724-5683.24.06546-3

75. Larsen TB, Skjøth F, Nielsen PB, Kjældgaard JN, Lip GYH. Comparative effectiveness and safety of non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study. *BMJ*. Published online June 16, 2016:i3189. doi:10.1136/bmj.i3189
76. Menichelli D, Del Sole F, Di Rocco A, et al. Real-world safety and efficacy of direct oral anticoagulants in atrial fibrillation: a systematic review and meta-analysis of 605 771 patients. *Eur Heart J Cardiovasc Pharmacother*. 2021;7(F11):f11-f19. doi:10.1093/ehjcvp/pvab002
77. Chen Y, Wu S, Ye J, et al. Predicting All-Cause Mortality Risk in Atrial Fibrillation Patients: A Novel LASSO-Cox Model Generated From a Prospective Dataset. *Front Cardiovasc Med*. 2021;8. doi:10.3389/fcvm.2021.730453
78. Lv H, Bi X, Shang S, et al. Machine learning model for predicting in-hospital cardiac mortality among atrial fibrillation patients. *Sci Rep*. 2025;15(1):29554. doi:10.1038/s41598-025-14579-8
79. Nguyen V, Mittal R. Machine Learning Prediction of ICU Mortality and Length of Stay in Atrial Fibrillation: A MIMIC-IV/MIMIC-III Study. *Healthcare*. 2026;14(3):356. doi:10.3390/healthcare14030356
80. Antonucci E, Poli D, Tosetto A, et al. The Italian START-Register on Anticoagulation with Focus on Atrial Fibrillation. Frati G, ed. *PLoS One*. 2015;10(5):e0124719. doi:10.1371/journal.pone.0124719
81. Bernardini A, Bindini L, Antonucci E, et al. Machine learning approach for prediction of outcomes in anticoagulated patients with atrial fibrillation. *Int J Cardiol*. 2024;407:132088. doi:10.1016/j.ijcard.2024.132088
82. Lu J, Bisson A, Bennamoun M, et al. Predicting multifaceted risks using machine learning in atrial fibrillation: insights from

GLORIA-AF study. *European Heart Journal - Digital Health*. 2024;5(3):235-246. doi:10.1093/ehjdh/ztae010

83. Huisman M V., Rothman KJ, Paquette M, et al. Antithrombotic Treatment Patterns in Patients with Newly Diagnosed Nonvalvular Atrial Fibrillation: The GLORIA-AF Registry, Phase II. *Am J Med*. 2015;128(12):1306-1313.e1. doi:10.1016/j.amjmed.2015.07.013
84. Ledziński Ł, Grześk E, Ledzińska M, Grześk G. Predicting Mortality in Atrial Fibrillation Patients Treated with Direct Oral Anticoagulants: A Machine Learning Study Based on the MIMIC-IV Database. *J Clin Med*. 2025;14(11):3697. doi:10.3390/jcm14113697
85. McGoon MD, Krichman A, Farber HW, et al. Design of the REVEAL Registry for US Patients With Pulmonary Arterial Hypertension. *Mayo Clin Proc*. 2008;83(8):923-931. doi:10.4065/83.8.923
86. Benza RL, Gomberg-Maitland M, Elliott CG, et al. Predicting Survival in Patients With Pulmonary Arterial Hypertension. *Chest*. 2019;156(2):323-337. doi:10.1016/j.chest.2019.02.004
87. Alabed S, Uthoff J, Zhou S, et al. Machine learning cardiac-MRI features predict mortality in newly diagnosed pulmonary arterial hypertension. *European Heart Journal - Digital Health*. 2022;3(2):265-275. doi:10.1093/ehjdh/ztac022
88. Hurdman J, Condliffe R, Elliot CA, et al. ASPIRE registry: Assessing the Spectrum of Pulmonary hypertension Identified at a REferral centre. *European Respiratory Journal*. 2012;39(4):945-955. doi:10.1183/09031936.00078411
89. Kheifets VO, Sweatt AJ, Gomberg-Maitland M, et al. Computational platform for doctor–artificial intelligence cooperation in pulmonary arterial hypertension prognostication: a pilot study. *ERJ Open Res*. 2023;9(1):00484-02022. doi:10.1183/23120541.00484-2022

90. Sweatt AJ, Hedlin HK, Balasubramanian V, et al. Discovery of Distinct Immune Phenotypes Using Machine Learning in Pulmonary Arterial Hypertension. *Circ Res*. 2019;124(6):904-919. doi:10.1161/CIRCRESAHA.118.313911
91. Hoeper MM, Kramer T, Pan Z, et al. Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. *European Respiratory Journal*. 2017;50(2):1700740. doi:10.1183/13993003.00740-2017
92. Strange G, Lau EM, Giannoulatou E, et al. Survival of Idiopathic Pulmonary Arterial Hypertension Patients in the Modern Era in Australia and New Zealand. *Heart Lung Circ*. 2018;27(11):1368-1375. doi:10.1016/j.hlc.2017.08.018
93. Suvon MNI, Tripathi PC, Alabed S, Swift AJ, Lu H. Multimodal Learning for Predicting Mortality in Patients with Pulmonary Arterial Hypertension. In: *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE; 2022:2704-2710. doi:10.1109/BIBM55620.2022.9995597
94. Attaripour Esfahani S, Baba Ali N, Farina JM, et al. A Comprehensive Review of Artificial Intelligence (AI) Applications in Pulmonary Hypertension (PH). *Medicina (B Aires)*. 2025;61(1):85. doi:10.3390/medicina61010085
95. Ledziński Ł, Grzešk G, Ziołkowski M, et al. Personalized Medicine in Pulmonary Arterial Hypertension: Utilizing Artificial Intelligence for Death Prevention. *J Clin Med*. 2025;14(23):8325. doi:10.3390/jcm14238325

8. Streszczenie

Wstęp: Współczesna medycyna stoi przed wyzwaniem zarządzania ogromnymi ilościami danych, których ilość rośnie wykładniczo. Tradycyjne metody analizy stają się niewystarczające, dlatego coraz większą rolę odgrywają zaawansowane narzędzia analityczne, takie jak sztuczna inteligencja (AI), uczenie maszynowe i duże modele językowe. AI umożliwia analizę złożonych danych, identyfikację wzorców i wspiera personalizację decyzji klinicznych. W pracy przedstawiono implementację metod uczenia maszynowego do predykcji śmiertelności w dwóch grupach pacjentów kardiologicznych: z migotaniem przedsionków (AF) leczonych doustnymi antykoagulantami niebędącymi antagonistami witaminy K oraz z tętniczym nadciśnieniem płucnym (PAH). W obu przypadkach skupiono się na identyfikacji i ocenie znaczenia wybranych parametrów klinicznych w przewidywaniu ryzyka zgonu, z wykorzystaniem nowoczesnych algorytmów sztucznej inteligencji.

Materialy i metody: Analizę przeprowadzono na dwóch dużych bazach danych: MIMIC-IV (6431 pacjentów z AF) oraz BNP-PL (1755 pacjentów z PAH). Przeprowadzono selekcję cech z użyciem regresji LASSO. Zbudowano modele klasyfikacyjne w oparciu o algorytmy wzmocnienia gradientowego, a ich wydajność oceniano na wydzielonych zbiorach testowych. Do interpretacji wyników i oceny istotności cech wykorzystano metodę SHAP.

Wyniki: Najlepszy model LightGBM dla pacjentów z AF osiągnął AUC 0.886, dokładność 0.862, czułość 0.913 i swoistość 0.859. W przypadku PAH model XGBoost uzyskał AUC 0.767, dokładność 0.738, czułość 0.800 i swoistość 0.733. W obu analizach metoda SHAP pozwoliła na wskazanie kluczowych czynników ryzyka, takich jak długość

hospitalizacji, czas pobytu na oddziale intensywnej opieki medycznej, obecność chorób współistniejących dla pacjentów z AF oraz dystans w teście 6-minutowego marszu, stężenie NT-proBNP, stosunek TAPSE/sPAP i wiek u pacjentów z PAH. Modele umożliwiły także indywidualną ocenę ryzyka dla poszczególnych pacjentów ze wskazaniem wpływu każdej z cech na predykcję.

Wnioski: Przeprowadzone badania potwierdzają, że modele uczenia maszynowego skutecznie wspierają predykcję śmiertelności w wybranych grupach pacjentów kardiologicznych, oferując narzędzia do bardziej spersonalizowanej opieki. Interpretowalność modeli zwiększa ich użyteczność kliniczną, umożliwiając identyfikację najważniejszych czynników ryzyka i wspierając proces podejmowania decyzji. Proponowane podejście stanowi cenne uzupełnienie tradycyjnych metod oceny ryzyka.

Słowa kluczowe: tętnicze nadciśnienie płucne, migotanie przedsionków, uczenia maszynowe, sztuczna inteligencja, kardiologia

9. Abstract

Introduction: Modern medicine faces the challenge of managing vast and exponentially growing amounts of data. Traditional analytical methods are becoming insufficient; therefore, advanced analytical tools such as artificial intelligence (AI), machine learning, and large language models are playing an increasingly important role. AI enables the analysis of complex data, identification of patterns, and supports the personalization of clinical decision-making. This study presents the implementation of machine learning methods for mortality prediction in two groups of cardiology patients: those with atrial fibrillation (AF) treated with non-vitamin K antagonist oral anticoagulants, and those with pulmonary arterial hypertension (PAH). In both cases, the focus was on identifying and assessing the importance of selected clinical parameters in predicting mortality risk, utilizing state-of-the-art AI algorithms.

Materials and Methods: The analysis was conducted on two large databases: MIMIC-IV (6,431 patients with AF) and BNP-PL (1,755 patients with PAH). Feature selection was performed using LASSO regression. Classification models were constructed based on gradient boosting algorithms, and their performance was evaluated on separate test sets. The SHAP method was used to interpret the results and assess feature importance.

Results: The best LightGBM model for AF patients achieved an AUC of 0.886, accuracy of 0.862, sensitivity of 0.913, and specificity of 0.859. For PAH, the XGBoost model obtained an AUC of 0.767, accuracy of 0.738, sensitivity of 0.800, and specificity of 0.733. In both analyses, the SHAP method enabled the identification of key risk factors, such as length of hospitalization, duration of stay in the intensive care unit, and presence

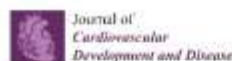
of comorbidities for AF patients, as well as distance in 6-minute walk test, plasma NT-proBNP concentration, TAPSE/sPAP ratio, and age for PAH patients. The models also allowed for individualized risk assessment for each patient, indicating the contribution of each feature to the prediction.

Conclusions: The conducted research confirms that machine learning models effectively support mortality prediction in selected groups of cardiology patients, providing tools for more personalized care. Model interpretability increases their clinical utility by enabling the identification of the most important risk factors and supporting the decision-making process. The proposed approach constitutes a valuable complement to traditional risk assessment methods.

Keywords: pulmonary arterial hypertension, atrial fibrillation, machine learning, artificial intelligence, cardiology

10. Załączniki

10.1. Publikacja I



Review

Artificial Intelligence Technologies in Cardiology

Lukasz Ledziński and Grzegorz Grzešek *

Department of Cardiology and Clinical Pharmacology, Faculty of Health Sciences, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, Uljskiego 75, 85-168 Bydgoszcz, Poland; 903349@doctorant.umk.pl

* Correspondence: g.grzesek@umk.pl

Abstract: As the world produces exabytes of data, there is a growing need to find new methods that are more suitable for dealing with complex datasets. Artificial intelligence (AI) has significant potential to impact the healthcare industry, which is already on the road to change with the digital transformation of vast quantities of information. The implementation of AI has already achieved success in the domains of molecular chemistry and drug discoveries. The reduction in costs and in the time needed for experiments to predict the pharmacological activities of new molecules is a milestone in science. These successful applications of AI algorithms provide hope for a revolution in healthcare systems. A significant part of artificial intelligence is machine learning (ML), of which there are three main types—supervised learning, unsupervised learning, and reinforcement learning. In this review, the full scope of the AI workflow is presented, with explanations of the most often-used ML algorithms and descriptions of performance metrics for both regression and classification. A brief introduction to explainable artificial intelligence (XAI) is provided, with examples of technologies that have developed for XAI. We review important AI implementations in cardiology for supervised, unsupervised, and reinforcement learning and natural language processing, emphasizing the used algorithm. Finally, we discuss the need to establish legal, ethical, and methodical requirements for the deployment of AI models in medicine.

Keywords: artificial intelligence; cardiology; machine learning



Citation: Ledziński, L.; Grzešek, G. Artificial Intelligence Technologies in Cardiology. *J. Cardiovasc. Dev. Dis.* **2023**, *10*, 302. <https://doi.org/10.3390/jcdd10060302>

Academic Editors: Ludmila Danilowska-Szymonowicz and Editada Welch

Received: 29 March 2023

Revised: 3 May 2023

Accepted: 4 May 2023

Published: 4 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There is a growing need to find a novel approach for using the enormous amounts of data that are collected every day for all aspects of life. According to the International Data Corporation (IDC), 153 exabytes of healthcare data were produced in 2013; it has been estimated that 2314 exabytes of healthcare data will have been produced in 2020 and that with this astronomical rate of growth, healthcare data will soon reach the zettabyte scale [1]. The healthcare industry has always generated huge quantities of information, and digital transformation has already begun. Information gathered from patients is often stored in electronic format—electronic health records (EHR)—and medical test results, imaging results, clinical notes, and in-patient interviews are being digitalized [2]. The volume of healthcare data is growing exponentially—not only data acquired by healthcare professionals (HCPs), but also data obtained by wearable sensors and via applications created to improve therapy adherence. This growing volume of data could be characterized as “4Vs”—high volume, high velocity, high veracity, and high variety—which makes data very difficult to manage using simple methods. Artificial intelligence (AI)—the computer of computers to mimic human intelligence—has significant potential to impact a complex domain such as medicine. AI is likely to be the basis for a significant revolution in healthcare systems. It has already had an immense impact on molecular chemistry and drug discovery. The techniques used in the prediction of molecules’ pharmacological activities are often based on artificial neural networks [3–5]. The possibility of using *in silico* experiments is crucial for achieving reductions in costs and time. Currently, there is considerable scientific

pressure to develop new drugs that will be more effective, applicable to new pathogens, and easier to administer. Furthermore, *in silico* experiments are much more in line with “green chemistry” than traditional experiments. There is even an established consortium, the Machine Learning for Pharmaceutical Discovery and Synthesis Consortium (MLPDSC), which includes the Massachusetts Institute of Technology (MIT) and representatives of 13 pharmaceutical and chemical companies [6–8].

As a discipline, AI includes systems that act like human intelligence—AI can learn and make decisions. Growing curiosity about artificial intelligence started in the 1950s when Alan Turing asked if machines could think and John McCarthy used the phrase “artificial intelligence” for the first time [9,10]. Now, thanks to improvements in hardware and affordability, interest in AI is increasing and it is being developed. In 2015, AlphaGo—a program based on a deep learning algorithm—defeated a human professional Go player for the first time in history [11].

The first big subpart of AI is machine learning (ML), which has the capability to develop systems that learn from retrospective data and achieve, as a result, classifications, clustering, or regression models. Algorithms of ML can detect patterns in complex data. Such a feature can be helpful in interpreting results and making personalized clinical decisions. Unlike traditional statistics, which are focused on creating scoring systems, machine learning develops automated clinical decision systems [12–18]. ML is already being used to create screening and diagnostic models [19–23].

Cardiology is a complex field of medicine that focuses on the diagnosis, treatment, and management of cardiovascular diseases. Due to its complexity, the diagnosis and treatment of cardiovascular diseases can be challenging. Cardiovascular diseases include a wide range of conditions that affect the heart and blood vessels, such as coronary artery disease, heart failure, arrhythmia, and valvular heart disease. These diseases can cause a range of symptoms, including chest pain, shortness of breath, fatigue, dizziness, and many deviations in laboratory parameters.

Medical imaging techniques, such as echocardiography, cardiac MRI, and CT scans, are among the primary diagnostic tools used in cardiology. These imaging methods provide detailed information about the heart's structure and function, enabling clinicians to detect abnormalities and diagnose cardiovascular diseases. Cardiac catheterization is another diagnostic procedure that is used to assess the heart's blood vessels and the blood flow to and from the heart.

The treatment of cardiovascular diseases varies, depending on the type and severity of the condition. Treatment options range from lifestyle modifications, such as diet and exercise, to medications, surgery, and medical devices such as pacemakers and implantable cardioverter-defibrillators (ICDs). Cardiac rehabilitation programs may also be recommended to help patients recover after a cardiac event and to improve their overall cardiovascular health.

Cardiology analysis differs from the analysis of other areas of the body, due to the heart's unique anatomy and physiology. The heart is a dynamic organ that constantly adapts to changes in the body's demand for oxygen and nutrients. The heart's intricate system of valves, chambers, and blood vessels requires specialized expertise to interpret and diagnose cardiovascular diseases accurately.

In recent years, AI technologies have been increasingly used in cardiology to improve the accuracy and efficiency of diagnosis, treatment, and management of cardiovascular diseases. AI-based image analysis algorithms can quickly and accurately detect abnormalities in medical images, and predictive models can analyze vast amounts of patient data to identify patterns and predict the likelihood of developing certain cardiovascular diseases. AI-based decision support systems can also assist clinicians in making treatment decisions, and wearable devices can monitor heart health and detect early signs of cardiovascular disease. With ongoing developments in AI technologies, the field of cardiology is poised to continue making significant advances in the diagnosis and management of cardiovascular diseases, ultimately leading to better patient outcomes.

The focus of this review is on the application of artificial intelligence technologies in the field of cardiology, in contrast to most of the similar reviews that concentrate on the use of AI in cardiology fields. The potential benefits of AI in cardiology are vast, including improved diagnostic accuracy, risk prediction, and treatment outcomes. Machine learning algorithms, neural networks, and natural language processing techniques are among the AI technologies that have been applied in cardiology. However, there are also challenges to overcome, including data privacy and security concerns, regulatory and legal issues, and the need for effective communication between AI systems and healthcare providers. The successful implementation and integration of AI into clinical practice require extensive testing and validation to ensure safety and effectiveness.

In the first part of the paper, general assumptions of AI are presented. Then, focus is placed on individual technologies and their potential applications in cardiology are described. Examples of works that use some of the aforementioned techniques are presented, followed by a discussion highlighting the challenges for the further implementation of artificial intelligence.

2. Introduction to AI in Healthcare

Obtaining a model by machine learning algorithms is preceded by the defining of the objective, data gathering, data preparation, data exploration, splitting data, and training using big data sets derived from many sources. The next step after training is tuning the hyperparameters, proceeding with testing, and validating.

The first step in the ML process is describing the problem to be faced. It is necessary to know what type of input data are available, which features are targeted, and of the nature of the problem. This step is followed by the most time-consuming and difficult element—gathering data from electronic health records and data curation by considering missing values, unnecessary information, etc. Exploratory data analysis is crucial in understanding the correlations, associations, patterns, and trends in the data.

Before beginning to build a model, it is essential to understand that there are three main types of ML: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning uses labeled data with a known outcome that is of interest for training set and for generating a model that adjusts weights until the model is fit for classification or regression. Classification analysis focuses on a matching class for each object; for example, in medicine, it can be used to check whether or not a patient has a disease. In a short term, the output of a classification model is usually nominal. On the other hand, regression models usually have continuous output that, for example, could be used in medicine to estimate the likelihood of disease. The learning process must be followed by cross-validation to ensure that the model is not underfitting or overfitting. An underfitting model will not be able to accurately predict or classify unseen data and training data; overfitting occurs when a model is ideal for training data but cannot perform accurately against unknown data [24]. See Figure 1. An artificial intelligence implementation workflow is presented on Figure 2.

Unsupervised learning uses unlabeled data as the input for analyzing and clustering by taking advantage of patterns that are present in the shared commonality of the data—a method that is widely used in medicine. Reinforcement learning (RL) might be described as the machine learning method for processing dynamic data, adapting an algorithm to change and optimize to achieve the best possible result. The learning is based on trial and error to achieve the maximum reward.

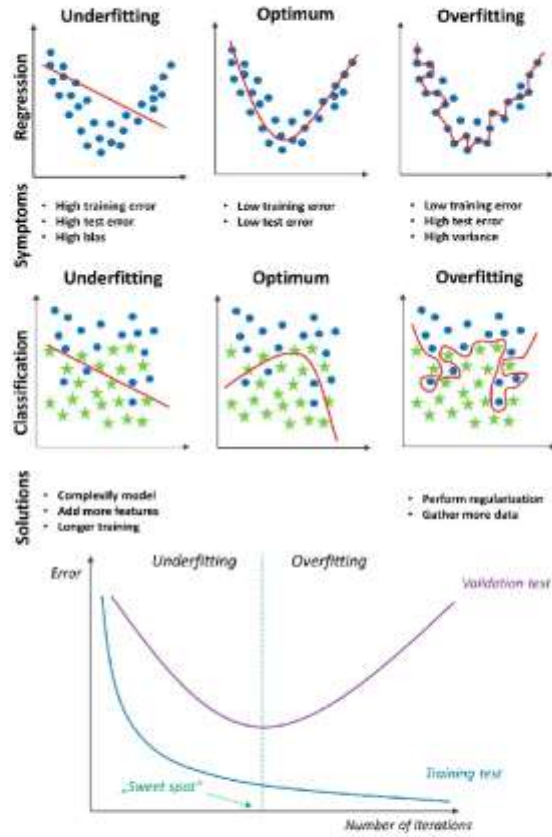


Figure 1. Underfitting and overfitting examples.

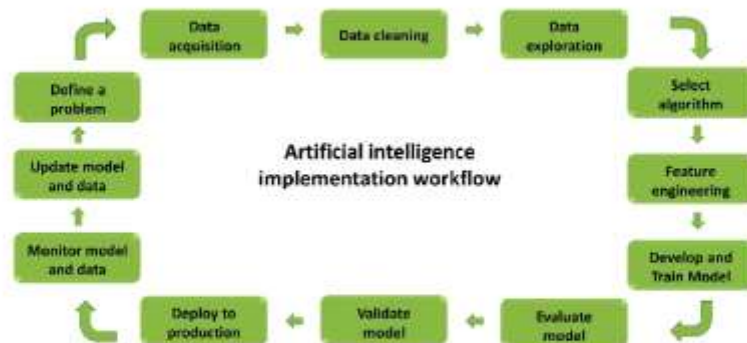


Figure 2. Artificial intelligence workflow.

3. The Most-Often-Used ML Algorithms

3.1. Supervised Learning

1. K-nearest neighbors (KNN)

KNN is a classification algorithm that is focused on finding similarity. The basic idea is to find a group of similar objects in a dataset and make predictions based on majority voting. KNN has two prominent hyperparameters to tune—distance function and K value. Determining the value of K is done by trial. The main issue is a scaling problem, which is usually solved by feature engineering.

2. Decision trees (DTs)

DTs are among the simplest models that are related to classification and regression. A decision tree is an algorithm that is presented with the tree structure, which splits data multiple times according to a specified threshold in the variables on the feature which, when split, provides the highest information gain. By division, different subsets of the dataset, called internal nodes, are created. Final subsets are called terminal leaves. An ensemble learning algorithm based on a decision tree is random forest (RF). In RF, many trees are created using random subsets of features and bootstrapped data; then, each tree votes by predicting a target class and all votes are tallied to reach the final prediction. The main advantage of RF is better accuracy without a higher computational workload. Another example of an ensemble decision-tree model is the gradient-boosted tree (GBT). Like RF, it can be used for solving a classification and regression problem by combining the outputs from individual trees. The main difference is in the way the individual trees are built and how the results are calculated. For RF purposes, bagging is implemented; on the other hand, GBT uses boosting, a technique that fits the sequence of models by adding higher weights to records, with higher residuals in each correct match. In other words, each new tree corrects the errors of the previous one, improving efficiency, accuracy, and interpretability. GBT has even higher accuracy than RF, but comes with a higher computational cost.

3. Support Vector Machines (SVMs)

The goal of SVMs is to delineate the hyperplanes that separate observations into two separate classes, with a maximum margin. SVMs were developed for linear classification, thanks to the development of kernel functions for nonlinear spaces. The choice of kernel functions must be based on existing data.

4. Naive Bayes (NB)

Bayes' rule provides a formula for determining the probability of Y by a given X . When functions are independent, the Bayes rule can be extended to the so-called naive Bayesian rule. Naive Bayes is a classification technique that uses the most basic knowledge of probability and makes the naive assumption that all characteristics are independent. An algorithm tries to find probabilities using known preliminary probabilities that are learned from training data. NB models are fast in computing and have worked quite well in solving problems such as spam filtering.

5. Artificial Neural Networks (ANNs)

Artificial neural networks simulate the behavior of the human brain and humanlike abilities to learn. The architecture of a neural network refers to the units, their activation functions, and the number of layers that may be found through experimentation. A single unit of an ANN is the neuron, which has an input value, weights and bias, a net sum, and an activation function. Inputs are multiplied by weights, summed along with a bias, then transformed by an activation function. Neural nets are trained by iterations with an optimization function. After each training cycle, called an epoch, an error metric is calculated based on the difference between the prediction and the goal. The derivatives of this error metric are computed and propagated back through the network using a technique called back-propagation. The factors (weights) of each neuron are then adjusted, depending on how much they contributed to the total error. This process is repeated iteratively until the network error falls below an acceptable threshold. Neurons are aggregates of layers—input layers—that collect initial data and send them further (each neuron from the input layer sends data to a neuron in the hidden layer), and hidden layers, which are intermediate states. Initially, the process of learning and searching for linear and nonlinear relationships is carried out. There may be multiple hidden layers (deep learning). The more hidden layers a network has, the more deeply it can find dependencies. Finally, the output layer (the result) is returned.

3.2. Unsupervised Learning

1. K-means

K-means algorithms are a prominent example of grouping algorithms. The five steps in a K-means algorithm can be distinguished. The first step is the selection of the number of clusters. Then, K starting points are drawn into the space (K first centroids). The third step is the mapping of all observations to the nearest centroid, using a given distance measure (the Euclidean distance), followed by finding a new centroid for each of the K -groups. The drawing and mapping steps are repeated until the stop condition is reached—for example, reaching the point where the assignment of observations to groups does not change, or reaching the assumed number of iterations. The main advantages of a K-means algorithm are simplicity and speed of operation. However, the disadvantages are the need to determine the number of groups, the sensitivity to the selection of the starting points, and the sensitivity to the impact of outliers.

2. Principal Component Analysis (PCA)

Principal component analysis is perhaps the most popular dimension-reduction algorithm. In a nutshell, it consists of projecting data to a space with a smaller number of dimensions in order to best preserve the data structure. It is primarily used to reduce the variables that describe a given phenomenon and to discover possible regularities among the features. A thorough analysis of principal components enables the identification of those initial variables that have a large impact on the appearance of individual principal components, i.e., those variables that make up a homogeneous group. It is worth remembering that, with PCA, we lose some information and we also lose interpretability.

3. Hierarchical clustering

Hierarchical clustering, as the name suggests, is an algorithm that builds a hierarchy of clusters. Hierarchical agglomeration clustering starts with all data points that are assigned to one's own cluster. Then, the two closest clusters are combined into the same cluster. The algorithm ends when there is only one remaining cluster. The algorithm merges groups in such a way that the variance inside them is as low as possible and the variance among the groups is as high as possible. These algorithms represent a detail-to-general approach. The approach implemented in the agglomeration algorithms works well when we decide to search for small groups (after several iterations, the process may be interrupted). The result of hierarchical clustering can be presented as a dendrogram. Hierarchical deglomeration clustering works the other way around, i.e., we start with one group and divide it into smaller groups in subsequent iterations.

4. A priori algorithms

A priori algorithms are well-known for mining association rules. They use frequent item sets for generating those rules and finding common patterns in a database. A priori algorithms are based on a "bottom-up" approach—i.e., frequent subsets are extended one at the time and then tested. The termination of the algorithm occurs when no further successful extension can be made. The main advantage of this algorithm is its ability to establish a relationship among events.

Healthcare AI applications are more likely to be based on supervised learning, while unsupervised learning is used for pre-processing dimension reductions, clustering, or finding relationships among observations [25,26]. Perhaps this usage of supervised learning in healthcare results from the enhanced explainability of supervised models [27]. There are high concerns about the safety, responsibility, and reliability of AI-based systems; a great deal of emphasis is placed on explainable artificial intelligence (XAI). A big warning that applies to a blind trust in AI was evident in the errors made by IBM Watson, a well-known algorithm used by many hospitals around the world, which recommended treatment for patients with cancer. As was pointed out by oncologists, the training dataset was unrealistic and too biased, and was also labeled by only few specialists instead of by guidelines or evidence-based medicine (EBM), which yielded many erroneous recommendations made by the algorithm [28,29]. It should be highlighted that medical malpractice is usually harmful for only a few patients, but a flawed AI model is a vast risk for a multitude of patients. XAI provides general information on how AI makes its decisions by disclosing the strengths and weaknesses of a program, why the model makes a specific decision, and specific criteria on the basis of which the program undertakes decisions. To overcome this explainability challenge, many novel technologies are being developed [30–32].

XAI employs a wide range of techniques to enhance the transparency and interpretability of machine learning algorithms. These techniques include:

1. **Rule-based systems:** Rule-based systems are among the earliest techniques used in XAI and they are still widely used today. These systems rely on predefined rules and decision trees that explicitly state how an algorithm makes decisions. This approach allows humans to trace the decision-making process, making it more transparent and understandable.
2. **Model interpretation techniques:** Model interpretation techniques are used to provide insight into how the machine learning model makes predictions. These techniques include visualization tools that help users understand the model's internal structure and feature importance analysis that shows which features are most important for the model's decisions.
3. **Local explanations:** Local explanations aim to provide an explanation for a single prediction. These techniques are used to explain why the machine learning algorithm made a specific decision for a given input. Examples of local explanation methods include LIME (local interpretable model-agnostic explanations) and SHapley additive exPlanations (SHAP).

4. Global explanations: Global explanations aim to provide explanations for the overall behavior of the machine learning model. These techniques include sensitivity analysis, which shows how changes in input features affect a model's output, and feature importance analysis, which shows which features are most important for a model's overall performance.
5. Human-in-the-loop: Human-in-the-loop (HITL) techniques involve incorporating human feedback into the machine learning process. These techniques allow humans to provide input and feedback at various stages of the model development process, increasing transparency and improving the model's performance.
6. GradCAM: gradient-weighted class activation mapping (GradCAM) is a visualization technique that highlights the regions of an image that are most important for the machine learning algorithm's decision-making process. GradCAM works by generating a heatmap that shows the areas of an image that had the highest activation in the convolutional layers of the deep neural network used by the machine learning model.

To compare a few models, it is necessary that some metrics show how good a model is in solving a problem. The most popular performance metrics are summarized in Tables 1 and 2.

Table 1. Description of performance metrics for classification.

Performance Metric	Description
False positive (FP)	Object incorrectly classified as positive
False negative (FN)	Object incorrectly classified as negative
True positive (TP)	Object correctly classified as positive
True negative (TN)	Object correctly classified as negative
Precision (PPV)	The fraction of TP among all positive classified
Sensitivity (TPR)/recall	The fraction of TP that were correctly classified
accuracy	The fraction of TP and TN that were correctly classified
F1 score	The harmonic mean of precision and recall
Specificity (TNR)	The fraction of TN that were correctly classified
Receiver operating characteristic curve	The curve between recall (Y-axis) and =false positive Rate = 1-specificity (X-axis)
Area under the curve ROC	Evaluates the overall quality of the model
Precision recall curve	The curve between precision (Y-axis) and recall (X-axis)
Precision recall AUC (PR-AUC)	Alternative for AUC-ROC based on the PR curve

Table 2. Description of performance metrics for regression.

Performance Metric	Description
Mean absolute error (MAE)	The mean of the absolute difference between the actual and predicted values in a dataset
Mean squared error (MSE)	The mean squared error between the predicted and actual values in a dataset
Root mean squared error (RMSE)	The square root of MSE
Coefficient of determination R^2	The proportion of variance explained by the model
Mean absolute percentage error (MAPE)	The mean of the absolute percentage errors of prediction

4. Natural Language Processing

An important feature of artificial intelligence, which is widely applied in medicine, is natural language processing (NLP). NLP is the capability of automatically analyzing and representing human language using computational methods. The variety of possibilities gained by applying NLP to medicine includes text classification, information extraction, and search engines for the effective use of clinical notes. Another aspect of natural language processing is the use of speech recognition and question-answering to develop chatbots

for patients [13]. Combining NLP and ML methods provides an opportunity to extract valuable resources from a large amount of unstructured EHR data that may be used in preparing clinical decision-making support systems, diagnostic tools, or recommendation systems based on search engines applied to scientific papers.

There are six main fazes of NLP:

1. Morphological analysis

Morphological analysis deals with understanding separate words according to their morphemes. We refer to the morpheme as the smallest unit of meaning.

2. Lexical analysis

Lexical analysis involves identifying and analyzing the structure of words and dividing an entire text into paragraphs, sentences, and words. To deal with lexical analysis, we often need to normalize lexicons.

3. Syntactic analysis

Syntax analysis is used to evaluate the compliance of a language with grammatical rules. Computer algorithms are used to apply grammatical rules to a group of words and to derive meanings from those words.

4. Semantic analysis

Semantic analysis is the use of computer algorithms to understand the meaning and interpretation of words and the structures of sentences. This is done by mapping syntactic structures and objects in the sentence domain. In other words, semantic analysis focuses on the interactions among meanings at the word level in a sentence. The best-known technique for semantic analysis is unit name recognition (NER).

5. Discourse integration

Discourse integration is concerned with how an immediately preceding sentence can influence the interpretation of the following sentence. Discourse integration focuses on the properties of the text, as a whole, that convey meaning by creating connections among sentences. In other words, it is focused on acquiring context from other sentences.

6. Pragmatic analysis

Pragmatic analysis explains how additional meaning is loaded into texts without actually being encoded into them. This requires a great deal of knowledge about the world, including an understanding of intentions, plans, and goals.

5. The implementation of Artificial Intelligence in Cardiology

5.1. Implementation of AI Technologies in Cardiology

Artificial intelligence has revolutionized various fields of medicine, including cardiology. With the advancements in technology, the use of AI in cardiology has become more sophisticated and its applications have widened. AI has the potential to improve the diagnosis, treatment management, and risk prediction of cardiac diseases, as well as the analysis of medical images such as echocardiograms or cardiac MRI scans.

Machine learning (ML) algorithms can be applied to clinical data, such as electrocardiograms (ECGs), echocardiograms, and medical imaging, to predict outcomes, stratify risk, and diagnose cardiovascular diseases. ML can also be used to identify patterns in data that may be invisible to the human eye, such as subtle changes in the heart's electrical activity, and to develop personalized treatment plans.

Deep learning (DL) is a subset of ML that has revolutionized the field of medical-imaging analysis. DL models can analyze large amounts of cardiovascular images, such as computed tomography (CT) scans and magnetic resonance imaging (MRI) scans, and can detect abnormalities with high accuracy. DL can also be used to create 3D reconstructions of the heart from multiple 2D images, enabling detailed analysis of the heart's structure and function.

Natural language processing technologies can be used to extract data from clinical documents such as electronic health records (EHRs) and physician notes, enabling the creation of large-scale clinical databases. NLP can also be used to develop algorithms that can identify and track specific cardiovascular risk factors, such as smoking, hypertension, and diabetes.

Computer-aided diagnosis (CAD) systems can be used to analyze medical images and provide diagnostic suggestions to clinicians. For example, CAD can be used to detect coronary artery disease by analyzing images of the coronary arteries and identifying areas of stenosis. CAD can also be used to analyze ECGs and identify abnormalities, such as arrhythmia.

Machine learning (ML) algorithms used in cardiology include the following:

1. The logistic regression algorithm is a type of regression analysis that is used for predicting the probability of a binary outcome. Logistic regression can be used in cardiology to predict the probability of a patient developing a cardiovascular disease based on various risk factors such as age, gender, and blood pressure.

Logistic regression models can be used to predict the likelihood of various cardiovascular events, including myocardial infarction (heart attack), stroke, and heart failure. Logistic regression models have been shown to be useful in identifying high-risk patients who may benefit from targeted interventions such as lifestyle modifications or medical therapy.

2. The decision trees algorithm is a machine learning method used for classification and regression analysis. In cardiology, decision trees can be used to create a model that can classify patients based on various cardiovascular risk factors, such as smoking status, blood pressure, and cholesterol levels.

Decision trees can be used to create decision support systems for clinicians to help them make more informed decisions about patient care. Decision trees can be used to identify patients who may be at high risk of developing cardiovascular disease and to develop personalized treatment plans for individual patients.

3. The random forest algorithm is an ensemble learning method that combines multiple decision trees to create a more accurate prediction model. In cardiology, random forests can be used to classify patients based on multiple decision trees, each tree using different combinations of risk factors.

Random forests can be used to create prediction models that are more accurate than individual decision trees. Random forests can identify the most important risk factors for cardiovascular disease and can help clinicians develop more effective treatment plans.

4. Support vector machines (SVMs) algorithms are supervised learning algorithms that are used for classification and regression analysis. In cardiology, SVMs can be used to identify patients at risk of developing cardiovascular disease by analyzing data such as age, blood pressure, and cholesterol levels.

SVM models can be used to identify patients at high risk of developing cardiovascular disease and to develop personalized treatment plans for these patients. SVM models can help clinicians make better-informed decisions about patient care and can improve patient outcomes.

Deep learning (DL) algorithms used in cardiology include the following:

1. Convolutional neural networks (CNNs) are a type of deep learning algorithm that is commonly used for image classification and recognition tasks. In cardiology, CNNs can analyze medical images such as CT scans, MRI scans, and echocardiograms, and identify various cardiovascular abnormalities such as heart disease and arrhythmia.

CNNs can be used to identify subtle changes in cardiac images that may indicate the presence of cardiovascular disease. CNNs can be used to develop more accurate and reliable diagnostic tools for clinicians, which can improve patient outcomes.

2. Recurrent neural networks (RNNs) are deep learning algorithms that are used for analyzing sequential data, such as time series. In cardiology, RNNs can analyze electrocardiogram (ECG) data and detect patterns and abnormalities that may indicate cardiovascular disease.

RNNs can be used to identify subtle changes in ECG data that may indicate the presence of cardiovascular disease. RNNs can be used to develop more accurate and reliable diagnostic tools for clinicians, which can improve patient outcomes.

Natural Language Processing (NLP) algorithms used in cardiology include the following:

1. Named entity recognition (NER) algorithms are NLP algorithms that are used to extract specific items such as diseases, symptoms, and medications from clinical documents such as electronic health records (EHRs). In cardiology, NER can be used to extract relevant information from patient records and help clinicians make better-informed decisions about patient care.

NER can be used to extract data from clinical documents such as EHRs and medical reports, which can be used to identify patterns and trends in patient data. NER can help clinicians identify patients who may be at high risk of developing cardiovascular disease and develop personalized treatment plans for these patients.

2. Sentiment analysis algorithms are NLP algorithms that are used to analyze text data and identify the sentiments or emotions conveyed by the text. In cardiology, sentiment analysis can be used to analyze patient feedback and reviews of cardiac treatments and procedures.

Sentiment analysis can be used to identify areas of patient care that may need improvement and to develop more patient-centered approaches to cardiac care. Sentiment analysis can help clinicians understand the patient experience and improve patient outcomes.

Data mining techniques used in cardiology include the following:

1. Association rule mining is a data mining technique that is used to identify patterns and relationships among different variables in a dataset. In cardiology, association rule mining can be used to identify risk factors and their relationships to various cardiovascular diseases.

Association rule mining can be used to identify complex patterns and relationships among different risk factors that may be missed by traditional statistical analysis. Association rule mining can help clinicians develop more effective risk prediction models and treatment plans for patients.

2. Clustering analysis is a data mining technique that is used to group similar data points together based on their characteristics. In cardiology, clustering analysis can be used to identify groups of patients with similar cardiovascular risk profiles.

Clustering analysis can be used to identify patient subgroups who may benefit from targeted interventions such as lifestyle modifications or medical therapy. Clustering analysis can help clinicians develop personalized treatment plans for individual patients and improve patient outcomes.

Computer-aided diagnosis (CAD) algorithms used in cardiology include the following:

1. Image segmentation is a type of algorithm used to separate an image into its component parts. In cardiology, image segmentation can be used to identify abnormalities in medical images, such as images of the heart and surrounding tissue.
2. Feature extraction is a type of algorithm used to identify specific features within an image. In cardiology, feature extraction can be used to identify specific features within medical images, such as the narrowing of coronary arteries.
3. Pattern recognition is a type of algorithm used to identify patterns in medical images such as the presence of plaques or blockages in arteries. In cardiology, pattern recognition can be used to identify patients who may be at risk of developing cardiovascular disease.

Examples of the usage of different AI technologies are briefly described in next section of this work.

5.2. Examples of Deployed Technologies

5.2.1. Supervised Learning

Supervised learning had been found to be useful for many applications in cardiovascular medicine [26,34–36]. Moghaddasi et al. [37] created an SVM-based model for the detection of the severity of mitral regurgitation by video analysis of 2D echocardiography with 99.38% sensitivity, 99.63% specificity, and 99.45% accuracy. Attia et al. [38] used paired 12-lead electrocardiogram and echocardiogram data from 44,959 patients to train a convolutional neural network for the identification of patients with ventricular dysfunction. They tested a created network on an independent dataset of 52,870 patients that yielded results of 86.3% sensitivity, 85.7% specificity, and 85.7% accuracy. Porumb et al. [39] proposed an innovative approach to predict the occurrence of hypoglycemia based on electrocardiography. They trained two deep learning models—a convolutional neural network (CNN) and a convolutional neural network combined with a recurrent neural network (CNN + RNN)—to show that a hypoglycemic event can be automatically detected using electrocardiography. The received results presented as follows: CNN model—81.7% sensitivity, 87.5% specificity, and 82.4% accuracy; CNN + RNN model—84.7% sensitivity, 84.5% specificity, and 85.7% accuracy. Echocardiography images were used for developing automatic measurements of left ventricular strain. Salte et al. created a pipeline of an artificial neural network that was able to estimate motion as an alternative to traditional speckle-tracking-based measures of strain, successfully classifying cardiac views and performing the timing of cardiac events [40]. Another application of CNN was proposed by Kusunose et al. [41], who trained a model for the detection of regional wall motion abnormalities in echocardiography view, which yielded an AUC similar to that of cardiologists (0.99 vs. 0.98).

Artificial neural networks have been successfully implemented for electrocardiogram interpretation. Their ability to find life-threatening arrhythmias provides for many applications of artificial neural networks [42]. Galloway et al. [43] reported the possibility of enabling the usage of a trained CNN model to screen hyperkalemia in patients with renal disease from electrocardiograms, achieving an AUC above 0.85. Decision trees were found to be helpful in discriminating between patients with pulmonary vein drivers and those with extra pulmonary vein drivers of atrial fibrillation based on electrocardiogram, to aid in the identification of patients with high acute success rates due to pulmonary vein isolation [44]. A comparison approach of AI application for the diagnosis of acute coronary syndrome was undertaken by Berikal et al. [45]. The parallel training of four different algorithms yielded an advantage of SVM over ANN, NB, and logistic regression, with respective accuracies of 99.13%, 90.10%, 88.75%, and 91.26%. A decision trees algorithm variation—LogitBoost—was used by Motwani et al. [46] for the prediction of mortality in patients with coronary artery disease. The proposed model outperformed the standard Framingham risk score (AUC 0.79 vs. AUC 0.61). Machine learning algorithms were shown to be better than conventional statistical models used in everyday clinical practice for the discrimination of readmission and mortality of heart-failure patients [47]. In addition, as Kakadiaris et al. [15] reported, an SVM-based trained model outperformed an American College of Cardiology/American Heart Association (ACC/AHA) risk calculator by overlooking fewer cardiovascular disease (CVD) events and recommending fewer drug therapies. ANN algorithms enable the identification of high-risk patients after a myocardial infarction. This allows for the preparation of personalized therapy [48]. A gradient boosting model was applied by Kogan et al. [49] for the classification of patients with pulmonary hypertension using EHR. AI can be used for identifying predictors of acute coronary syndrome events or for risk stratification and the diagnosis of pulmonary arterial hypertension [13,16,50–54].

5.2.2. Unsupervised Learning

Unsupervised learning was mainly used for clustering and grouping and as a pre-processing method for dimensional reduction. Karwath et al. [55] proposed the implementation of hierarchical clustering for distinguishing prognostic responses for β -blocker therapy in patients with heart failure and reduced the left ventricular ejection fraction. Cikes et al. [56] suggested that unsupervised learning algorithms can be used to combine standard clinical parameters and echocardiographic images to achieve an interpretable measure for clinicians' classification of a phenotypical heterogeneous heart failure cohort and to identify patients who are more likely to respond to treatment. Unsupervised ML is still underrated in comparison to supervised ML, but it has been found to be an application in the analysis of EHR or genetic data for the automatization of data extraction [57,58]. An unsupervised algorithm called topological data analysis, used on combined EHR and genetic data, allowed Li et al. to reveal the presence of three separate subtypes of diabetes type 2 [59].

5.2.3. Reinforcement Learning

The reinforcement learning state-action-reward-state-action (SARSA) algorithm performed a selection of dofetilide dose adjustments based on a negative reward for unsuccessful initiation [60]. Ghesu et al. [61] presented a novel method for real-time anatomical landmarks detection that was high in performance and robustness. Unfortunately, there are still only a few applications of reinforcement learning in cardiology. A possible niche for RL is the personalization of therapy for certain patient characteristics, because of the algorithm's inherent decision-making construction.

5.2.4. Natural Language Processing

NLP is a pre-processing method that makes native text understandable for machines. There are many nonmedical applications of natural language processing e.g., chat-boxes and search engines. There is hope for more studies using NLP in the future, due to previously demonstrated improvement in algorithm performance with unstructured data [62–68]. Adzal et al. [69] recommended NLP as a tool for the rapid and efficient testing of peripheral arterial diseases based on clinical narrative notes. The NLP model outperformed billing code algorithms, yielding 91.2% sensitivity, 92.5% specificity, and 91.8% accuracy. Another example of improvement achieved by deploying NLP was reported by Ashburner et al. [70], whose study demonstrated that merging clinical and demographic features with incorporating narrative data from the EHR can greatly improve the efficiency of a model. Deploying NLP to studies may reduce misclassifications caused by the extraction of additional information available only in EHR.

Table 3 summarizes most of mentioned works and shows the main used algorithm. According to research by Jiang et al., two of the most often-used AI algorithms for images, genetic data, and electrophysiology are SVMs and ANNs [71]. The wide application of SVMs in cardiovascular studies is dictated by their quite simple explainability of results and their ease of implementation. On the other hand, ANNs are usually very complex and they make it difficult to explain how a result was achieved; nevertheless, they have prominent computation power to handle image or electrophysiological data.

Table 3. Summarized studies and used algorithms.

Author	Study	Algorithm	Performance (Accuracy/AUROC/Precision)
Koçan et al. [49]	A machine learning approach to identifying patients with pulmonary hypertension using real-world electronic health records	XGBoost	0.92 AUROC
Mohammad et al. [48]	Development and validation of an artificial neural network algorithm to predict mortality and admission to hospital for heart failure after myocardial infarction: a nationwide population-based study	ANN	0.85–0.78 AUROC
Moghaddasi et al. [37]	Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos	SVM	99.45%
Atis et al. [38]	Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram	CNN	85.70%
Porumb et al. [39]	Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG	CNN, RNN, Grad-CAM	82.40–85.70%
Salle et al. [40]	Artificial intelligence for automatic measurement of left ventricular strain in echocardiography	ANN	97–98%
Kusunose et al. [41]	A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images	CNN	0.99–0.97 AUROC
Bertkol et al. [45]	Diagnosis of acute coronary syndrome with a support vector machine	SVM, ANN, NB, Logistic Regression	90.10–99.13%
Motwani et al. [46]	Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis	DT—Logitboost	0.79 AUROC
Kakadiaris et al. [15]	Machine learning outperforms ACC/AHA CVD risk calculator in MESA	SVM	0.92 AUROC
Karwar et al. [50]	Risk stratification in pulmonary arterial hypertension using Bayesian analysis	Bayesian Network	0.80 AUROC
Galloway et al. [43]	Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram	CNN	0.85–0.88 AUROC
Luongo et al. [44]	Machine learning enables noninvasive prediction of atrial fibrillation driver location and acute pulmonary vein ablation success using the 12-lead ECG	DT	78.26%
Karwath et al. [53]	Redefining β -blocker response in heart failure patients with sinus rhythm and atrial fibrillation: a machine learning cluster analysis	Hierarchical clustering, Variational autoencoders (VAEs), K-means	N/A
Cikci et al. [56]	Machine learning-based phenotyping in heart failure to identify responders to cardiac resynchronization therapy	K-means, Multiple Kernel Learning	N/A

Table 3. Cont.

Author	Study	Algorithm	Performance (Accuracy/AUROC/Precision)
Li et al. [59]	Identification of type 2 diabetes subgroups through topological analysis of patient similarity	Topological data analysis	N/A
Ghesu et al. [61]	Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans	Deep reinforcement learning	Accuracy improved by 20–30%
Levy et al. [60]	Applications of machine learning in decision analysis for dose management for doxorubicin	PCA, K-means, reinforcement learning—SARSA	86–93%
Garvin et al. [62]	Automating quality measures for heart failure using natural language processing: a descriptive study in the department of veterans affairs	NLP	Precision 98.7%
Shah et al. [63]	Impact of different electronic cohort definitions to identify patients with atrial fibrillation from the electronic medical record	NLP	0.89 AUROC
Kaspar et al. [64]	Underestimated prevalence of heart failure in hospital inpatients: a comparison of ICD codes and discharge letter information	NLP	Precision 96%
Patel et al. [65]	Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records	NLP	Precision 96%
Mahajan et al. [66]	Combining structured and unstructured data for predicting risk of readmission for heart failure patients	NLP	0.63 AUROC
Galper et al. [67]	Comparison of adverse event and device problem rates for transcatheter aortic valve replacement and mitralclip procedures as reported by the transcatheter valve therapy registry and the Food and Drug Administration postmarket surveillance data	NLP	N/A
Afzal et al. [68]	Mining peripheral arterial disease cases from narrative clinical notes using natural language processing	NLP	91.8%
Ashburner et al. [70]	Natural language processing to improve prediction of incident atrial fibrillation using electronic health records	NLP	N/A

6. Discussion

Artificial intelligence is around us in every aspect of our lives. Smartwatches are no longer extraordinary and in addition to their many other features they can monitor health. Wearable technology can not only check how many calories have been burned or one's actual heartbeat rate, they are also capable of detecting atrial fibrillation by applying an AI algorithm to differentiate the heart's rhythm and provide recommendations to patients [72–76].

There is an urgent need to create official international requirements—legal, ethical, and methodical—for applying AI models to medicine [77–79]. Due to the rapid increase in the number of scientific articles about AI, some guidelines have already appeared, but many important questions remain unsettled [80–82]. Today, with growing computing power of super computers, we are able to analyze and process terabytes of data. The outcomes of such analyses should be used by physicians who can judge and reassess acquired information and, then, make decisions based on their knowledge, their experience, and, finally, AI suggestions. Machine learning might be able to overcome some of the limitations

that occur in still-used statistical-based models. The challenges that must be overcome for the widespread use of AI in medicine can be divided into two groups: algorithm/model-related challenges and data-related challenges. The most important algorithm-related problems are standardization, reproducibility, explainability of predicted results, and legal and ethical responsibility for the outcome consequences for patients. Wrong assumptions and incorrect model-type selection may lead to faulty results and could be fatal for patients. The WHO has established ethical guidance for the implementation of artificial intelligence in healthcare [83].

The main challenges related to data involve the method of acquisition and safety. A highlighted case involving the transfer of patient-identifiable EHR without permission, from the Royal Free London NHS Foundation Trust to Google Deep Mind for algorithm development, has presented a major concern for public trust [84]. European regulations on data protection and privacy are well established after the publication of the General Data Protection Regulation by the European Union in 2016 [85].

Another tremendous challenge is related to data quality. The expression “garbage in, garbage out” precisely describes how poorly gathered and selected data can lead to erroneous assumptions and unsuccessful model predictions. It is crucial to develop international EHR registries that will not be outdated, biased, or discriminatory. A good EHR database should characterize information with robustness, transparency, trust, and verifiability [86]. A solid and frequently checked security system must also be in place before full AI implementation [36]. The safety issue is not only related to data storing, but also to hacking resistance. The hackability of implanted medical devices is a life-threatening problem because of possible fatal consequences for patients. Recent research revealed the vulnerability of AI healthcare systems, including FDA warnings about the vulnerability of Medtronic insulin pumps to cyberattack [87–89].

The development of novel technologies is not only needed to benefit patients by improving diagnostic accuracy and providing personalized therapy that is focused on extending the quantity and quality of life. According to a Stanford University medicine report, there is an urgent need to help physicians with their work because of a high percentage of burnout, resulting in serious medical errors, inefficient patient care, and higher costs of care [90]. One of the reasons behind clinicians’ burnout is the vast amount of their work that is focused on preparing EHR—the time spent in preparing EHR is more than double the amount of time spent in front of patients. The loss of productivity was estimated to range from USD 90 billion to USD 140 billion [91]. Facilitating a clinical-decision process for physicians is essential, especially in the treatment of rare and lethal diseases. A great example is the future utilization of artificial intelligence to predict a pharmacotherapy response for pulmonary hypertension patients. See Figure 3. Such an approach could possibly improve the length and quality of life via personalized treatment. Utilizing a combination of AI and wearable technology to predict and manage cardiovascular therapy has the potential to usher in an era of precision medicine [92]. Undoubtedly, emerging machine learning approaches such as “Federated Learning”, which uses decentralized data and allows models to be trained on vastly distributed datasets based on patients across different clinics, will enable the training of highly accurate models for diagnosing cardiovascular diseases, while ensuring patient data privacy. However, this approach requires considerable further development [93].

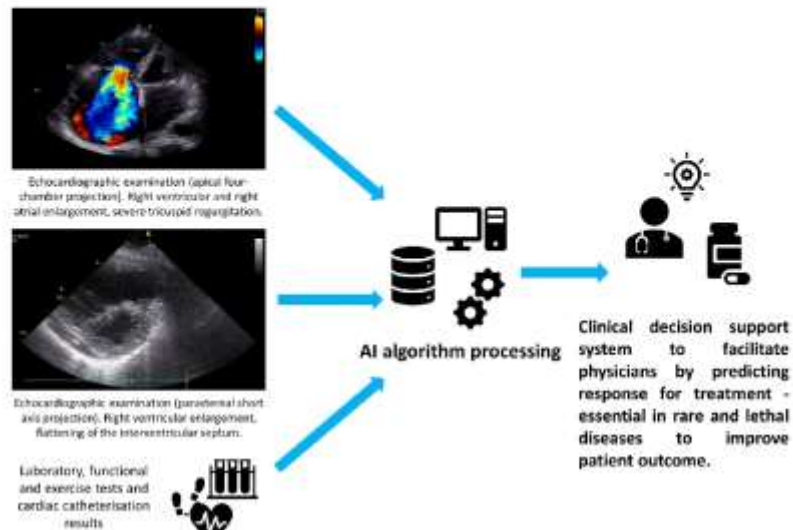


Figure 3. Future of clinical decision support systems.

7. Conclusions

Making use of artificial intelligence is crucial for future technological improvements. Medicine is facing increasingly new challenges and a global crisis, in the face of which machine learning has become a perfect complement to traditional forms of care and assistance. AI has many advantages that could be useful and helpful for physicians and patients. It also has some disadvantages and pitfalls, which everyone using AI-based technology should be aware of [94]. It is clear that artificial intelligence has the power to revolutionize the field of cardiology, offering new and innovative ways to diagnose, predict, and treat cardiovascular diseases. By leveraging AI technologies, we can improve the accuracy and efficiency of diagnosis, develop personalized treatment plans, and, ultimately, improve patient outcomes. While there are undoubtedly challenges, such as patient privacy, bias, and discrimination, that need to be addressed to ensure the responsible and ethical use of AI in cardiology, these concerns can be mitigated through the establishment of clear ethical guidelines and standards. By harnessing the power of AI to analyze complex data sets and identify patterns and trends, we have the potential to achieve more precise and effective cardiovascular care. Ultimately, the use of AI in cardiology offers a unique opportunity to transform the way we approach patient care and promote equitable healthcare outcomes, paving the way for a future of more personalized and effective cardiovascular medicine. In the words of Stephen Hawking, “Our future is a race between the growing power of technology and the wisdom with which we use it.”

Author Contributions: Conceptualization, E.L. and G.G.; writing—original draft preparation, E.L.; writing—review and editing, G.G.; visualization, E.L.; supervision, G.G.; project administration, G.G.; funding acquisition, G.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AI	Artificial intelligence
ANN	Artificial neural network
AUC	Area under the curve
CNN	Convolutional neural network
DT	Decision tree
EHR	Electronic health records
FDA	Food and Drug Administration
GBI	Gradient-boosted tree
GradCAM	Gradient-weighted class activation mapping
HCP	Healthcare professionals
IDC	International Data Corporation
KNN	K-Nearest neighbor
ML	Machine learning
MLPDS	Machine Learning for Pharmaceutical Discovery and Synthesis Consortium
NB	Naïve Bayes
NER	Unit name recognition
NLP	Natural language processing
PCA	Principal component analysis
RF	Random Forest
RL	Reinforcement learning
RNN	Recurrent neural network
SARSA	State-action-reward-state-action
SHAP	SHapley Additive exPlanations
SVM	Support vector machines
VAEs	Variational autoencoders
WHO	World Health Organization
XAI	Explainable artificial intelligence

References

- Reinsel, D.; Gantz, J.; Rydning, J. *The Digitization of the World from Edge to Core*; International Data Corporation: Framingham, MA, USA, 2018.
- Stanford Medicine 2017 Health Trends Report *Harnessing the Power of Data in Health*; Stanford Medicine: Stanford, CA, USA, 2017.
- Badura, A.; Krysiński, J.; Nowaczyk, A.; Bucirski, A. Application of Artificial Neural Networks to Prediction of New Substances with Antimicrobial Activity against *Escherichia Coli*. *J. Appl. Microbiol.* **2021**, *130*, 40–49. [\[CrossRef\]](#)
- Badura, A.; Krysiński, J.; Nowaczyk, A.; Bucirski, A. Prediction of the Antimicrobial Activity of Quaternary Ammonium Salts against *Staphylococcus Aureus* Using Artificial Neural Networks. *Antib. J. Chem.* **2021**, *14*, 103233. [\[CrossRef\]](#)
- Lin, X.; Li, X.; Lin, X. A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules* **2020**, *25*, 1375. [\[CrossRef\]](#)
- Struble, T.J.; Alvarez, J.C.; Brown, S.P.; Chytil, M.; Cisar, J.; Desjardis, R.L.; Engkvist, O.; Frank, S.A.; Greve, D.R.; Griffin, D.J.; et al. Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis. *J. Med. Chem.* **2020**, *63*, 8682. [\[CrossRef\]](#)
- ASKCOS Homepage. Available online: <https://askcos.mit.edu/> (accessed on 9 October 2022).
- MLPDS—Machine Learning for Pharmaceutical Discovery and Synthesis Consortium. Available online: <https://mlpds.mit.edu/> (accessed on 9 October 2022).
- McCarthy, J.; Minsky, M.L.; Rochester, N.; Shannon, C.E. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Mag.* **1955**, *27*, 12. [\[CrossRef\]](#)
- Turing, A.M.L. —Computing Machinery and Intelligence. *Mind* **1950**, *11X*, 435–460. [\[CrossRef\]](#)
- Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. Mastering the Game of Go without Human Knowledge. *Nature* **2017**, *550*, 354–359. [\[CrossRef\]](#) [\[PubMed\]](#)

12. Fox, K.A.A.; Dabbous, O.H.; Goldberg, R.J.; Pieper, K.S.; Eagle, K.A.; van de Werf, F.; Avezum, A.; Goodman, S.C.; Flather, M.D.; Anderson, E.A.; et al. Prediction of Risk of Death and Myocardial Infarction in the Six Months after Presentation with Acute Coronary Syndrome: Prospective Multinational Observational Study (GRACE). *BMJ* **2006**, *333*, 1091. [\[CrossRef\]](#)
13. Benza, R.L.; Gomburg-Maitland, M.; Miller, D.P.; Frost, A.; Frantz, R.P.; Foreman, A.J.; Badesch, D.B.; McGoon, M.D. The REVEAL Registry Risk Score Calculator in Patients Newly Diagnosed with Pulmonary Arterial Hypertension. *Chest* **2012**, *141*, 354–362. [\[CrossRef\]](#)
14. Kaniwa, M.; Raina, A.; Lohmüller, L.; Kraissangkarn, J.; Benza, R. The Use of Risk Assessment Tools and Prognostic Scores in Managing Patients with Pulmonary Arterial Hypertension. *Curr. Hypertens. Rep.* **2019**, *21*, 45. [\[CrossRef\]](#)
15. Kakadiaris, I.A.; Vrigkas, M.; Yen, A.A.; Kuznetsova, T.; Budoff, M.; Naghavi, M. Machine Learning Outperforms ACC/AHA CVD Risk Calculator in MESA. *J. Am. Heart Assoc.* **2018**, *7*, e009476. [\[CrossRef\]](#) [\[PubMed\]](#)
16. D'Ascenzo, F.; de Filippo, O.; Gallone, G.; Mittone, G.; Deriu, M.A.; Iannaccone, M.; Ariza-Solé, A.; Liebetrau, C.; Manzano-Fernández, S.; Quadri, G.; et al. Machine Learning-Based Prediction of Adverse Events Following an Acute Coronary Syndrome (PRAISE): A Modelling Study of Pooled Datasets. *Lancet* **2021**, *397*, 199–207. [\[CrossRef\]](#)
17. Alas, A.M.; Bolton, T.J.; Angelantonio, F.; Rudd, J.H.F.; van der Schaar, M. Cardiovascular Disease Risk Prediction Using Automated Machine Learning: A Prospective Study of 423,604 UK Biobank Participants. *PLoS ONE* **2019**, *14*, e0213653. [\[CrossRef\]](#) [\[PubMed\]](#)
18. de Baquer, D.; Ueda, P.; Reiner, Z.; de Sauter, J.; de Smedt, D.; Luvic, D.; Gotcheva, N.; Pras, Z.; Pogosova, N.; Mirakhimov, E.; et al. Prediction of Recurrent Event in Patients with Coronary Heart Disease: The EUROASPIRE Risk Model Results from a Prospective Study in 27 Countries in the WHO European Region—The EURObservational Research Programme (EORP) of the European Society of Cardiology (ESC). *Eur. J. Prev. Cardiol.* **2022**, *29*, 328–339. [\[CrossRef\]](#)
19. Hippisley-Cox, J.; Coupland, C.; Brindle, P. Development and Validation of QRESK3 Risk Prediction Algorithms to Estimate Future Risk of Cardiovascular Disease: Prospective Cohort Study. *BMJ* **2017**, *357*, j2099. [\[CrossRef\]](#)
20. Piesko, K.; Hiczkiewicz, J.; Budzianowski, P.; Rzeźniczak, J.; Budzianowski, J.; Ilaszczyński, J.; Słowinski, R.; Burchardt, P. Machine-Learned Models Using Hematological Inflammation Markers in the Prediction of Short-Term Acute Coronary Syndrome Outcomes. *J. Transl. Med.* **2018**, *16*, 334. [\[CrossRef\]](#)
21. Piesko, K.; Hiczkiewicz, J.; Budzianowski, P.; Budzianowski, J.; Rzeźniczak, J.; Piesko, K.; Burchardt, P. Predicting Long-Term Mortality after Acute Coronary Syndrome Using Machine Learning Techniques and Hematological Markers. *Dis. Markers* **2019**, *2019*, 9056402. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Brojer, N.; Cozzi, B.; Gao, M.; Nichols, M.; Revzin, M.; Balu, S.; Fuoloma, J.; Bae, J.; Setji, N.; Hernandez, A.; et al. Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission. *JAMA Netw. Open* **2020**, *3*, e1920733. [\[CrossRef\]](#)
23. Kwon, J.M.; Kim, K.H.; Akkus, Z.; Jeon, K.H.; Park, J.; Oh, B.J. Artificial Intelligence for Detecting Mitral Regurgitation Using Electrocardiography. *J. Electrocardiol.* **2020**, *50*, 151–157. [\[CrossRef\]](#)
24. What Is Overfitting? IBM. Available online: <https://www.ibm.com/cloud/learn/> (accessed on 10 October 2022).
25. Mathur, P.; Srivastava, S.; Xu, X.; Mehta, J.L. Artificial Intelligence, Machine Learning, and Cardiovascular Disease. *Clin. Med. Insights Cardiol.* **2020**, *14*, 1179546820927404. [\[CrossRef\]](#)
26. Johnson, K.W.; Torres Souza, J.; Glicksberg, B.S.; Shamsir, K.; Miotto, R.; Ali, M.; Ashley, E.; Dudley, J.T. Artificial Intelligence in Cardiology. *J. Am. Coll. Cardiol.* **2018**, *71*, 2668–2679. [\[CrossRef\]](#)
27. Amann, T.; Blasième, A.; Vayena, E.; Frey, D.; Madai, V.I. Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 310. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Ross, C.; Swednitz, L.; Hogan, A. IBM's Watson Supercomputer Recommended "Unsafe and Incorrect" Cancer Treatments, Internal Documents Show. *Stat News*, 2018. Available online: <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/> (accessed on 5 May 2023).
29. Topol, E.J. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nat. Med.* **2019**, *25*, 44–56. [\[CrossRef\]](#)
30. Wang, K.; Tian, J.; Zheng, C.; Yang, H.; Ren, J.; Liu, Y.; Han, Q.; Zhang, Y. Interpretable Prediction of 3-Year All-Cause Mortality in Patients with Heart Failure Caused by Coronary Heart Disease Based on Machine Learning and SHAP. *Circul. Int. Med.* **2021**, *137*, 104813. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Neves, I.; Folgado, D.; Santos, S.; Berandas, M.; Campagner, A.; Ronzio, L.; Cabrita, F.; Gambos, H. Interpretable Heartbeat Classification Using Local Model-Agnostic Explanations on ECGs. *Comput. Biol. Med.* **2021**, *133*, 104393. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Knapik, S.; Malhi, A.; Saluja, R.; Trümbling, K. Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 740–770. [\[CrossRef\]](#)
33. Laranjo, L.; Dunn, A.G.; Long, H.L.; Kocoballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.S.; et al. Conversational Agents in Healthcare: A Systematic Review. *J. Am. Med. Assoc.* **2018**, *320*, 1248. [\[CrossRef\]](#)
34. Krittanawong, C.; Zhang, H.J.; Wang, Z.; Aydar, M.; Kitab, T. Artificial Intelligence in Precision Cardiovascular Medicine. *J. Am. Coll. Cardiol.* **2017**, *69*, 2657–2664. [\[CrossRef\]](#)
35. Ichhaporia, D. Artificial Intelligence in Cardiology. *Trends Cardiovasc. Med.* **2022**, *32*, 34–41. [\[CrossRef\]](#)

36. Al'Anaf, S.J.; Anchouche, K.; Singh, G.; Slomka, P.J.; Kolli, K.K.; Kumar, A.; Pandey, M.; Mafakal, G.; van Rosendaal, A.R.; Beccy, A.N.; et al. Clinical Applications of Machine Learning in Cardiovascular Disease and Its Relevance to Cardiac Imaging. *Eur. Heart J.* **2019**, *40*, 1975–1986. [\[CrossRef\]](#)
37. Moghaddas, H.; Nourian, S. Automatic Assessment of Mitral Regurgitation Severity Based on Extensive Textural Features on 2D Echocardiography Videos. *Comput. Biol. Med.* **2016**, *73*, 47–55. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Attia, Z.I.; Kapa, S.; Lopez-Jamenez, E.; McKie, P.M.; Ludwig, D.J.; Satam, G.; Pellikka, P.A.; Enriquez-Sarano, M.; Noseworthy, P.A.; Munger, T.M.; et al. Screening for Cardiac Contractile Dysfunction Using an Artificial Intelligence-Enabled Electrocardiogram. *Nat. Med.* **2019**, *25*, 70–74. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Fonamb, M.; Stranges, S.; Pescapé, A.; Picchia, L. Precision Medicine and Artificial Intelligence: A Pilot Study on Deep Learning for Hypoglycemic Events Detection Based on ECG. *Sci. Rep.* **2020**, *10*, 170. [\[CrossRef\]](#)
40. Salte, I.M.; Østvik, A.; Smestad, E.; Melichova, D.; Nguyen, T.M.; Karlsson, S.; Brannvand, H.; Haugaa, K.H.; Edvardsen, T.; Lovstakken, L.; et al. Artificial Intelligence for Automatic Measurement of Left Ventricular Strain in Echocardiography. *JACC Cardiovasc. Imaging* **2021**, *14*, 1918–1928. [\[CrossRef\]](#)
41. Kusuruse, K.; Abe, T.; Hago, A.; Fukuda, D.; Yamada, H.; Harada, M.; Sato, M. A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images. *JACC Cardiovasc. Imaging* **2020**, *13*, 374–381. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Feeney, A.K.; Chung, M.K.; Madabhushi, A.; Attia, Z.I.; Cikos, M.; Firooznia, M.; Friedman, P.A.; Kalscheur, M.M.; Kapa, S.; Narayan, S.M.; et al. Artificial Intelligence and Machine Learning in Arrhythmias and Cardiac Electrophysiology. *Circ. Arrhythmia Electrophysiol.* **2020**, *13*, E007952. [\[CrossRef\]](#)
43. Galloway, C.D.; Valys, A.V.; Shrestha, J.B.; Treiman, D.L.; Peterson, F.L.; Gundotra, V.J.; Albert, D.E.; Attia, Z.I.; Carter, R.E.; Asirvatham, S.J.; et al. Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia from the Electrocardiogram. *JAMA Cardiol.* **2019**, *4*, 428–436. [\[CrossRef\]](#)
44. Luongo, G.; Azzioli, L.; Schuler, S.; Brocchi, M.W.; Almeida, T.P.; Martinez, J.P.; Soriano, D.C.; Luik, A.; Müller-Eberhard, B.; Jafar, A.; et al. Machine Learning Enables Noninvasive Prediction of Atrial Fibrillation Driver Location and Acute Pulmonary Vein Ablation Success Using the 12-Lead ECG. *Cardiovasc. Digit. Health J.* **2021**, *2*, 126–136. [\[CrossRef\]](#)
45. Berikol, G.B.; Yildiz, O.; Özcan, T. Diagnosis of Acute Coronary Syndrome with a Support Vector Machine. *J. Med. Syst.* **2016**, *40*, 84. [\[CrossRef\]](#)
46. Motcari, M.; Dey, D.; Berman, D.S.; Germano, G.; Achenbach, S.; Al-Mallah, M.H.; Andreini, D.; Badoff, M.J.; Cademartini, F.; Callister, T.Q.; et al. Machine Learning for Prediction of All-Cause Mortality in Patients with Suspected Coronary Artery Disease: A 5-Year Multicentre Prospective Registry Analysis. *Eur. Heart J.* **2017**, *38*, 500. [\[CrossRef\]](#)
47. Shin, S.; Austin, P.C.; Ross, H.J.; Abdel-Qadir, H.; Freitas, C.; Tomlinson, G.; Chocco, D.; Mahendran, M.; Lawler, P.R.; Billia, E.; et al. Machine Learning vs. Conventional Statistical Models for Predicting Heart Failure Readmission and Mortality. *ESC Heart Fail* **2021**, *8*, 106–115. [\[CrossRef\]](#)
48. Mohammad, M.A.; Olsen, K.K.W.; Koo, S.; Gale, C.P.; Rylance, R.; Jernberg, T.; Baron, T.; Hpaik, J.; James, S.; Lindahl, B.; et al. Development and Validation of an Artificial Neural Network Algorithm to Predict Mortality and Admission to Hospital for Heart Failure after Myocardial Infarction: A Nationwide Population-Based Study. *Lancet Digit. Health* **2022**, *4*, e37–e45. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Kogut, E.; Diddis, E.M.; Lee, E.; Nnewiwe, A.; Stamatiadis, D.; Mataraso, S.; Quinn, D.; Rosenberg, D.; Chehoud, C.; Bridges, C. A Machine Learning Approach to Identifying Patients with Pulmonary Hypertension Using Real-World Electronic Health Records. *Int. J. Cardiol.* **2023**, *374*, 95–99. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Kanwar, M.K.; Gombert, M.; Hooper, M.; Pausch, C.; Pitow, D.; Strange, G.; Anderson, J.J.; Zhao, C.; Scott, J.V.; Druzdzal, M.J.; et al. Risk Stratification in Pulmonary Arterial Hypertension Using Bayesian Analysis. *Eur. Respir. J.* **2020**, *56*, 2000008. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Zhu, F.; Xu, D.; Liu, Y.; Lou, K.; He, Z.; Zhang, H.; Sheng, Y.; Yang, R.; Li, X.; Kong, X.; et al. Machine Learning for the Diagnosis of Pulmonary Hypertension. *Kardiologia* **2020**, *60*, 96–101. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Kwon, J.M.; Kim, K.H.; Medina-Inojosa, J.; Jeon, K.H.; Park, J.; Oh, B.H. Artificial Intelligence for Early Prediction of Pulmonary Hypertension Using Electrocardiography. *J. Heart Lung Transplant.* **2020**, *39*, 805–814. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Leht, A.; Hellensamp, K.; Unsöld, B.; Mushemi-Bake, S.; Shah, A.M.; Hasenfull, G.; Seidler, T. A Machine Learning Approach for the Prediction of Pulmonary Hypertension. *PLoS ONE* **2019**, *14*, e0224453. [\[CrossRef\]](#)
54. Kiely, D.G.; Doyle, O.; Drage, E.; Jenner, H.; Salvatelli, V.; Daniels, F.A.; Rigg, J.; Schmitt, C.; Samyashkin, Y.; Lawrie, A.; et al. Utilising Artificial Intelligence to Determine Patients at Risk of a Rare Disease: Idiopathic Pulmonary Arterial Hypertension. *Paediatr. Clin.* **2019**, *9*, 2045894019880549. [\[CrossRef\]](#)
55. Karwarth, A.; Bunting, K.v.; Gill, S.K.; Tica, O.; Pendleton, S.; Aziz, F.; Barsky, A.D.; Chemburumong, S.; Duan, J.; Mobley, A.R.; et al. Redefining β -Blocker Response in Heart Failure Patients with Sinus Rhythm and Atrial Fibrillation: A Machine Learning Cluster Analysis. *Lancet* **2021**, *398*, 1427–1435. [\[CrossRef\]](#)
56. Cikos, M.; Sanchez-Martinez, S.; Claggett, B.; Duchateau, N.; Piella, G.; Butakoff, C.; Catherine Pouleur, A.; Knappe, D.; Iffertsen, T.; Kutylka, V.; et al. Machine Learning-Based Phenotyping in Heart Failure to Identify Responders to Cardiac Resynchronization Therapy. *Eur. J. Heart Fail.* **2019**, *21*, 74–85. [\[CrossRef\]](#)

57. Miotto, R.; Li, L.; Kidd, B.A.; Dudley, J.T. *Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records (OPEN)*; Nature Publishing Group: Berlin, Germany, 2016. [\[CrossRef\]](#)
58. Shomorony, I.; Cirulli, E.T.; Huang, L.; Napier, L.A.; Heister, R.R.; Hicks, M.; Cohen, I.; Yu, H.-C.; Swisher, C.L.; Schenker-Ahmed, N.M.; et al. An Unsupervised Learning Approach to Identify Novel Signatures of Health and Disease from Multimodal Data. *Genome Med.* **2020**, *12*, 7. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Li, L.; Cheng, W.-Y.; Glicksberg, B.S.; Gottesman, O.; Tamler, R.; Chen, R.; Bottinger, E.P.; Dudley, J.T. Identification of Type 2 Diabetes Subgroups through Topological Analysis of Patient Similarity. *Sci. Transl. Med.* **2015**, *7*, 311ra174. [\[CrossRef\]](#)
60. Levy, A.E.; Biswas, M.; Weber, R.; Tarakji, K.; Chung, M.; Noseworthy, P.A.; Newton-Cheh, C.; Rosenberglid, M.A. Applications of Machine Learning in Decision Analysis for Dose Management for Doxetidine. *PLoS ONE* **2019**, *14*, e0227324. [\[CrossRef\]](#)
61. Ghisu, F.C.; Georgescu, B.; Zheng, Y.; Gebic, S.; Maior, A.; Hornegger, J.; Comaniciu, D. Multi-Scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 176–189. [\[CrossRef\]](#)
62. Carvin, J.H.; Kim, Y.; Gibbel, G.T.; Matheny, M.E.; Rodd, A.; Bray, B.E.; Heidenreich, P.; Bolton, D.; Heuvrland, J.; Kelly, N.; et al. Automating Quality Measures for Heart Failure Using Natural Language Processing: A Descriptive Study in the Department of Veterans Affairs. *JMIR Med. Inform.* **2018**, *6*, e5. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Shah, R.U.; Mukherjee, R.; Zhang, Y.; Jones, A.E.; Springer, J.; Hackett, I.; Steinberg, R.A.; Lloyd-Jones, D.M.; Chapman, W.W. Impact of Different Electronic Cohort Definitions to Identify Patients With Atrial Fibrillation From the Electronic Medical Record. *J. Am. Heart Assoc.* **2020**, *9*, e014527. [\[CrossRef\]](#)
64. Kasper, M.; Fette, G.; Güdel, G.; Seidlmeier, L.; Ertl, M.; Dietrich, G.; Greger, H.; Puppe, F.; Störk, S. Underestimated Prevalence of Heart Failure in Hospital Inpatients: A Comparison of ICD Codes and Discharge Letter Information. *Clin. Res. Cardiol.* **2018**, *107*, 778–787. [\[CrossRef\]](#)
65. Patel, Y.R.; Robbins, J.M.; Kurgansky, K.E.; Inzun, T.; Orkaby, A.R.; McLean, R.R.; Ho, Y.L.; Chn, K.; Michael Gaziano, J.; Djousse, L.; et al. Development and Validation of a Heart Failure with Preserved Ejection Fraction Cohort Using Electronic Medical Records. *BMC Cardiovasc. Disord.* **2018**, *18*, 128. [\[CrossRef\]](#)
66. Mahajan, S.M.; Ghani, R. Combining Structured and Unstructured Data for Predicting Risk of Readmission for Heart Failure Patients. *Stud. Health Technol. Inform.* **2019**, *264*, 238–242. [\[CrossRef\]](#)
67. Galper, B.Z.; Beery, D.B.; Leighton, G.; Englander, L.L. Comparison of Adverse Event and Device Problem Rates for Transcatheter Aortic Valve Replacement and Mitralclip Procedures as Reported by the Transcatheter Valve Therapy Registry and the Food and Drug Administration Postmarket Surveillance Data. *Am. Heart J.* **2018**, *198*, 64–74. [\[CrossRef\]](#)
68. Reading, Turchio, M.; Volodarsky, A.; Pathak, J.; Wright, D.N.; Tchong, J.E.; Slowinski, D. Systematic Review of Current Natural Language Processing Methods and Applications in Cardiology. *Heart* **2022**, *108*, 909–916. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Afzal, N.; Sohn, S.; Abrams, S.; Scott, C.G.; Chaudhry, R.; Liu, H.; Kullo, L.J.; Arzuda-Olsen, A.M. Mining Peripheral Arterial Disease Cases from Narrative Clinical Notes Using Natural Language Processing. *J. Vasc. Surg.* **2017**, *65*, 1753–1761. [\[CrossRef\]](#)
70. Ashburner, J.M.; Chang, Y.; Wang, X.; Khurshid, S.; Anderson, C.D.; Dahal, K.; Weisenfeld, D.; Cai, T.; Liao, K.P.; Waghlikar, K.B.; et al. Natural Language Processing to Improve Prediction of Incident Atrial Fibrillation Using Electronic Health Records. *J. Am. Heart Assoc. Cardiovasc. Cerebrovasc. Dis.* **2022**, *11*, 26014. [\[CrossRef\]](#)
71. Jiang, F.; Jiang, Y.; Zhi, H.; Deng, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. Artificial Intelligence in Healthcare: Past, Present and Future. *Stroke Vasc. Neurol.* **2017**, *2*, 230–243. [\[CrossRef\]](#)
72. Perez, M.V.; Mahaffey, K.W.; Heitlin, H.; Runnels, J.S.; Garcia, A.; Ferris, T.; Balasubramanian, V.; Russo, A.M.; Rajmane, A.; Cheung, L.; et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. *N. Engl. J. Med.* **2019**, *381*, 1909–1917. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Wasserlauf, J.; You, C.; Patel, R.; Valys, A.; Albert, D.; Passerun, R. Smartwatch Performance for the Detection and Quantification of Atrial Fibrillation. *Circ. Arrhythmia Electrophysiol.* **2019**, *12*, e008834. [\[CrossRef\]](#)
74. Raja, J.M.; Elsaye, C.; Roman, S.; Carr, B.; Four-Ghaz, I.; Nanda, A.; Maturana, M.; Khouram, R.N. Apple Watch, Wearables, and Heart Rhythm: Where Do We Stand? *Ann. Transl. Med.* **2019**, *7*, 417. [\[CrossRef\]](#)
75. Seshadri, D.R.; Bittel, B.; Browsky, D.; Houghtaling, P.; Drummond, C.K.; Desai, M.Y.; Gillinov, A.M. Accuracy of Apple Watch for Detection of Atrial Fibrillation Circulation. *Circulation* **2020**, *141*, 702–703. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Watson, X.; D'Souza, J.; Cooper, D.; Markham, R. Artificial Intelligence in Cardiology: Fundamentals and Applications. *Intern. Med. J.* **2022**, *52*, 912–920. [\[CrossRef\]](#)
77. Sengupta, P.P.; Shrestha, S.; Beribon, B.; Messes, E.; Donal, E.; Tison, G.H.; Min, J.K.; D'hooge, J.; Voigt, J.-U.; Dudley, J.; et al. Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): A Checklist. *JACC Cardiovasc. Imaging* **2020**, *13*, 2017–2035. [\[CrossRef\]](#)
78. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G.M. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Br. J. Surg.* **2015**, *102*, 148–158. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Moons, K.G.M.; Altman, D.G.; Reitsma, J.B.; Ioannidis, J.P.A.; Macaskill, P.; Sneyers, E.W.; Vickers, A.J.; Ransohoff, D.F.; Collins, G.S. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann. Intern. Med.* **2015**, *162*, W1–W73. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Quer, G.; Arnaud, R.; Henne, M.; Arnaud, R. Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review. *J. Am. Coll. Cardiol.* **2021**, *77*, 300–313. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Schwalbe, N.; Wahl, B. Artificial Intelligence and the Future of Global Health. *Lancet* **2020**, *395*, 1579. [\[CrossRef\]](#)

82. Westerlund, A.M.; Hawe, J.S.; Heinig, M.; Schunkert, H. Risk Prediction of Cardiovascular Events by Exploration of Molecular Data with Explainable Artificial Intelligence. *Int. J. Mol. Sci.* **2021**, *22*, 10291. [\[CrossRef\]](#)
83. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health*; World Health Organization: Geneva, Switzerland, 2021; ISBN 9789240029200.
84. Morley, J.; Taddeo, M.; Floridi, L. Google Health and the NHS: Overcoming the Trust Deficit. *Lancet Digit. Health* **2019**, *1*, e389. [\[CrossRef\]](#)
85. General Data Protection Regulation (GDPR). Compliance Guidelines. Available online: <https://gdpr.eu/> (accessed on 15 October 2022).
86. Lorkowski, J.; Grzegorzewska, O.; Pokorski, M. Artificial Intelligence in the Healthcare System: An Overview. In *Best Practice in Health Care*; Pokorski, M., Ed.; Advances in Experimental Medicine and Biology; Springer: Cham, Switzerland, 2021; Volume 1339, pp. 1–10. [\[CrossRef\]](#)
87. FDA Warns Patients and Health Care Providers about Potential Cybersecurity Concerns with Certain Medtronic Insulin Pumps | FDA. Available online: <https://www.fda.gov/news-events/press-announcements/fda-warns-patients-and-health-care-providers-about-potential-cybersecurity-concerns-certain> (accessed on 15 October 2022).
88. Tobissam, A.; Sofi, Z.; Alkhater, W.; Shikda, A. Cybersecurity Issues in Implanted Medical Devices. In Proceedings of the 2018 International Conference on Computer and Applications, ICCA 2018, Beirut, Lebanon, 25–26 August 2018; pp. 110–115. [\[CrossRef\]](#)
89. Finlayson, S.G.; Chung, H.W.; Kohane, I.S.; Beam, A.L. Adversarial Attacks Against Medical Deep Learning Systems. *arXiv* **2018**, arXiv:1804.05296. [\[CrossRef\]](#)
90. Stanford Medicine. *2020 Health Trends Report: the Rise of the Data-Driven Physician*; Stanford Medicine: Stanford, CA, USA, 2020.
91. Lin, S. The Present and Future of Team Documentation: The Role of Patients, Families, and Artificial Intelligence. *Mayo Clin. Proc.* **2020**, *95*, 852–858. [\[CrossRef\]](#)
92. Visco, V.; Izzu, C.; Mancusi, C.; Rispoli, A.; Todeschi, M.; Virtuoso, N.; Giano, A.; Gioia, R.; Melfi, A.; Sorio, B.; et al. Artificial Intelligence in Hypertension Management: An Ace up Your Sleeve. *J. Cardiovasc. Dis. Prev.* **2023**, *10*, 74. [\[CrossRef\]](#)
93. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Reviewing Federated Machine Learning and Its Use in Diseases Prediction. *Sensors* **2023**, *23*, 2112. [\[CrossRef\]](#) [\[PubMed\]](#)
94. Scott, J.; Carter, S.; Colera, E. Clinician Checklist for Assessing Suitability of Machine Learning Applications in Healthcare. *BMJ Health Care Inform.* **2021**, *28*, 100251. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

10.2. Publikacja II

Article

Predicting Mortality in Atrial Fibrillation Patients Treated with Direct Oral Anticoagulants: A Machine Learning Study Based on the MIMIC-IV Database

Lukasz Ledziński ¹, Elżbieta Grzešek ^{2,*}, Małgorzata Ledzińska ³ and Grzegorz Grzešek ¹

¹ Department of Cardiology and Clinical Pharmacology, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, 85-168 Bydgoszcz, Poland; S03345@doctorint.umk.pl (L.L.); g.grzesek@umk.pl (G.G.)

² Department of Pediatrics, Hematology and Oncology, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, 85-104 Bydgoszcz, Poland

³ Department of Rheumatology and Connective Tissue Diseases, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, 85-168 Bydgoszcz, Poland; malgorzata.ledzinska.90@gmail.com

* Correspondence: elag@cm.umk.pl

Abstract: **Background/Objectives:** Atrial fibrillation (AF) is a common arrhythmia linked to increased mortality and significant healthcare burden, especially in the elderly. Direct Oral Anticoagulants (DOACs) are crucial for stroke prevention in AF, offering benefits over traditional vitamin K antagonists. Despite scoring systems like HATCH and CHA2DS2-VASc, their predictive ability for mortality in AF patients is limited. This study aims to use machine learning to predict mortality within six months of hospital discharge in AF patients treated with DOACs. **Methods:** Using the MIMIC-IV database, data from 6431 AF patients were analyzed. Feature selection was done with the LASSO algorithm. Five machine learning models were built: Logistic Regression, Random Forest, XGBoost, LightGBM, and AdaBoost, using 27 features. The top two models were tested on a separate dataset. SHAP values explained model predictions and feature importance. **Results:** The best model, LightGBM, achieved an AUC of 0.886, accuracy of 0.862, sensitivity of 0.913, and specificity of 0.839. SHAP values highlighted the importance of length of hospital stay, ICU duration, and comorbidities. The model's interpretability allows for identifying individual patient risk factors, applicable in clinical practice. **Conclusions:** This study demonstrates that machine learning models effectively predict mortality in AF patients treated with DOACs, potentially enhancing personalized patient care.

Keywords: machine learning; atrial fibrillation; direct oral anticoagulants; cardiology



Academic Editor: Carlos Escobedo

Received: 26 April 2023

Revised: 26 May 2023

Accepted: 22 May 2023

Published: 28 May 2023

Citation: Ledziński, L.; Grzešek, E.; Ledzińska, M.; Grzešek, G. Predicting Mortality in Atrial Fibrillation Patients Treated with Direct Oral Anticoagulants: A Machine Learning Study Based on the MIMIC-IV Database. *J. Clin. Med.* **2023**, *12*, 3697. <https://doi.org/10.3390/jcm12123697>

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Atrial fibrillation (AF) is a common supraventricular arrhythmia associated with an increased risk of mortality. It is characterized by a rapid and chaotic heart rhythm, resulting in disrupted blood flow and potentially leading to serious complications such as stroke and heart failure, which adversely affect patient longevity and quality of life [1–3]. The high prevalence of AF imposes a significant burden on the healthcare system [4,5]. Additionally, the aging population significantly increases the incidence of this condition [5,6]. Direct Oral Anticoagulants (DOACs) have become a cornerstone in the management of atrial fibrillation (AF), particularly for stroke prevention [7]. Unlike traditional vitamin K antagonists (VKAs) such as warfarin, DOACs offer several advantages, including predictable pharmacokinetics, fewer dietary restrictions, and a lower risk of intracranial hemorrhage [8].

DOACs, which include agents such as dabigatran, rivaroxaban, apixaban, and edoxaban, directly inhibit specific clotting factors (thrombin or factor Xa), providing effective anticoagulation with a more favorable safety profile. Clinical trials and real-world studies have demonstrated that DOACs are at least as effective as VKAs in preventing stroke and systemic embolism in patients with non-valvular AF, with a reduced risk of major bleeding [9]. Consequently, current guidelines recommend DOACs as the preferred anticoagulant therapy for most patients with AF, barring specific contraindications. Artificial Intelligence has found widespread application in various fields, including medicine, and particularly in cardiology, where its capabilities are widely used [10,11]. Currently used scoring systems such as HATCH, HAVOC, and CHA2DS2-VASc have shown limited ability to predict mortality in patients with atrial fibrillation (AF), as demonstrated in several studies [12–14]. Attempts to use machine learning for predicting mortality in patients with atrial fibrillation (AF) have been undertaken by various researchers [15–17]. Given the increasing number of patients with AF and the widespread use of DOACs in this patient group, we decided to utilize machine learning tools to predict mortality within six months of hospital discharge in this population.

2. Materials and Methods

The study focuses on utilizing retrospective data from the Medical Information Mart for Intensive Care (MIMIC)-IV version 3.1 database [18,19]. This is a large, anonymized database from the Beth Israel Deaconess Medical Center in Boston, MA, containing data for 364,627 patients from 2008 to 2022. Access to the MIMIC database was obtained through a multi-step process to ensure ethical and responsible use of the data. First, researchers completed the Collaborative Institutional Training Initiative (CITI) “Data or Specimens Only Research” course, which provides essential training on the ethical considerations and responsibilities associated with handling sensitive health data. Following this, a Data Use Agreement (DUA) was signed and submitted, outlining the terms and conditions for using the MIMIC data in compliance with all applicable regulations. Finally, credentialing was obtained through PhysioNet, the platform hosting the MIMIC database [20]. Upon successful completion of these steps, access to the MIMIC database was granted for research purposes. Patients with atrial fibrillation (AF) ($n = 32,026$) were selected for the study, identified by ICD codes [‘42731’, ‘4891’, ‘148’, ‘1480’, ‘1481’, ‘14811’, ‘14819’, ‘1482’, ‘14820’, ‘14821’, ‘1489’], who initiated treatment with DOACs ($n = 6431$) (dabigatran/rivaroxaban/apixaban/edoxaban) during hospitalization. Patients who died during their hospital stay were excluded from the cohort. Information was selected on medical history, hospital stay, Charlson Comorbidity Index (CCI) score, demographic details such as age, gender, race, marital status, hospital and ICU stay, administered medications, microbiology test results, performed procedures, diagnostic codes, and laboratory results, resulting in a total of 600 variables.

Data cleaning was conducted by removing all variables with more than 10% missing values. Feature engineering was performed to extract information on the total ICU length of stay, the total number of diagnoses, hospital length of stay, the reason for hospital admission (whether it was related to a procedure or observation), the absence of performed procedures, negative microbiology test results, heparin treatment, and readmission within 60 days of discharge. Patient age was initially divided into 10-year intervals and then encoded in ascending order. Features such as ‘Gender’, ‘Congestive Heart Failure’, ‘Dementia’, ‘Diabetes without Complications’, ‘Paraplegia’, ‘Metastatic Solid Tumor’, ‘Readmission within 60 days’, ‘Marital Status: Widowed’, ‘Observation Admission’, ‘Surgical Admission’, ‘Microbiology: Negative’, ‘No applied procedures’, ‘Heparin’, and ‘Obesity’ were encoded as binary variables (0/1).

The remaining features were continuous variables: 'Anion Gap', 'Bicarbonate', 'Hemoglobin', 'Potassium', 'Sodium', 'Urea Nitrogen', 'White Blood Cells', 'Creatinine', 'Sum of ICU Length of Stay', 'Number of Diagnoses', 'Charlson Comorbidity Index', and 'Hospital Days'. All continuous variables were standardized prior to the application of Logistic Regression (LR) and the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm. Patient records containing missing values were excluded from the dataset. Missing data are a common challenge in clinical datasets and can impact the validity and generalizability of predictive models. Several imputation strategies are available, including simple approaches such as mean or median substitution, as well as more sophisticated methods like k-nearest neighbors (KNN) imputation and MICE. Simple imputation methods are computationally efficient but may underestimate variability and distort associations between variables. KNN imputation leverages the similarity between observations, potentially preserving local data structure, but can be sensitive to outliers and the choice of k. MICE, an advanced iterative technique, models each variable with missing values as a function of other variables, generating multiple plausible datasets and thus accounting for imputation uncertainty. In this study, we explored multiple imputation using MICE in addition to listwise deletion. However, the application of MICE did not result in a meaningful improvement in model performance or alter the main findings. Given the lack of demonstrable benefit from imputation, we opted for listwise deletion to maximize methodological transparency and interpretability while minimizing the risk of introducing imputation-related bias. This approach is further supported by the absence of evidence for systematic differences between complete and incomplete cases: Only data from the initial hospitalization when atrial fibrillation (AF) was first diagnosed were considered. The class distribution was 6086:345. The prepared dataset was then split into training and testing sets in an 80:20 ratio. Feature selection was performed using the LASSO algorithm, resulting in the selection of 27 features for model construction. LASSO is a regularization technique that enhances both the prediction accuracy and interpretability of regression models by imposing a constraint on the absolute values of the model parameters, effectively shrinking some coefficients to zero. This method is particularly advantageous in high-dimensional datasets, where it aids in mitigating overfitting and improving model generalization by selecting a subset of the most informative features. Consequently, LASSO is widely employed for feature selection, as it systematically identifies and retains only the most relevant predictors, thereby simplifying the model and enhancing its performance. A machine learning model was developed using the eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), Light Gradient-Boosting Machine (LightGBM), Random Forest (RF), and LR algorithms to predict patient mortality from any cause within six months of hospital discharge. XGBoost is an advanced gradient-boosting algorithm that enhances performance and speed through optimized implementation, making it highly effective for structured data analysis. AdaBoost is an ensemble learning technique that improves classification accuracy by iteratively adjusting the weights of weak classifiers, thereby focusing on previously misclassified instances. LightGBM is a gradient-boosting framework that leverages tree-based learning algorithms, optimized for high efficiency and scalability, particularly suitable for large-scale data processing. Random Forest is an ensemble method that constructs multiple decision trees during training and aggregates their predictions, known for its robustness and ability to reduce overfitting. Logistic Regression is a statistical method that models the probability of a binary outcome using a logistic function, valued for its simplicity and interpretability in binary classification tasks.

During training, internal validation was performed using RepeatedStratifiedKFold ($n_splits = 5$, $n_repeats = 5$) to ensure the accuracy and robustness of the model. Hyperparameter tuning for all models was conducted using the HyperOpt library [21]. HyperOpt

is a powerful optimization library in Python designed for hyperparameter tuning, which uses Bayesian optimization and other search algorithms to efficiently explore the hyperparameter space. It is particularly useful for improving model performance by finding the optimal set of hyperparameters in a computationally efficient manner. The top two models with the best internal validation performance were used to make predictions on the test set.

To assess the model's quality, standard metrics for binary classification were utilized, such as accuracy, specificity, sensitivity, the Matthews correlation coefficient, and area under the ROC curve.

Accuracy denotes the proportion of correctly classified instances among all instances. It reflects the overall performance of the model and is straightforward to interpret. However, in the context of imbalanced datasets, accuracy can be deceptive as it may not accurately represent the model's performance on the minority class.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN). \quad (1)$$

Specificity, or the true negative rate, measures the model's ability to correctly identify negative instances. It is crucial in situations where minimizing Type I errors (false positives) is essential. High specificity ensures that the majority of actual negative cases are accurately identified by the model.

$$\text{Specificity} = TN / (TN + FP). \quad (2)$$

Sensitivity, also known as recall, assesses the model's ability to correctly identify positive instances. It is vital in scenarios where minimizing Type II errors (false negatives) is important. High sensitivity ensures that most actual positive cases are detected by the model.

$$\text{Sensitivity} = TP / (TP + FN). \quad (3)$$

The Matthews Correlation Coefficient (MCC) is a balanced measure of binary classification performance that considers all four values of the confusion matrix. It is particularly useful for imbalanced datasets. The MCC ranges from -1 to $+1$, where $+1$ indicates perfect prediction, 0 indicates random prediction, and -1 indicates total disagreement between prediction and observation.

$$\text{MCC} = ((TP \cdot TN) - (FP \cdot FN)) / \sqrt{((TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN))}. \quad (4)$$

The Receiver Operating Characteristic—Area Under the Curve (ROC-AUC) represents the area under the curve that illustrates the relationship between sensitivity and 1-specificity across various decision thresholds. It evaluates the model's discriminative ability; a higher value indicates a better capability to distinguish between classes:

- True Positive (TP): Correctly predicted death of patient;
- True Negative (TN): Correctly predicted survival of patient;
- False Positive (FP): Incorrectly predicted survival of patient, type I error;
- False Negative (FN): Incorrectly predicted death of patient, type II error.

All computations presented in this study were conducted in Python 3.10.12 on a CPU featuring 24 threads at 4.5 GHz and 64 GB of RAM. Packages versions: pandas (2.1.4), numpy (1.26.4), hyperopt (0.2.7), shap (0.46.0), scikit-learn (1.4.2), xgboost (2.1.3), lightgbm (4.5.0).

3. Results

3.1. The General Characterization of Patients Group

The study population consisted of patients with atrial fibrillation (AF) who were treated with Direct Oral Anticoagulants (DOACs) identified from the MIMIC-IV database, with baseline characteristics compared between survivors and non-survivors. The baseline characteristics of patients were compared between the group of survivors and non-survivors. Categorical variables were expressed as count and frequency, compared using the chi-square test. Continuous variables were presented as mean $\pm$ standard deviation (SD) or median and first and third quartiles (Q1, Q3). Continuous variables were analyzed using Mann–Whitney U test. A two-side test result with p -value < 0.05 was considered statistically significant. Table 1, below, summarizes the statistics of continuous and categorical variables used to build the model.

Table 1. Statistical analysis of 27 features used for building model.

Feature	Survivors (n = 6086)	Non-Survivors (n = 345)	p-Value (Two-Sided)
Age [years], median (Q1, Q3)	76 (68,84)	81 (73,88)	0.00000
Gender, male, n (%)	3416 (56.13%)	167 (48.41%)	0.00589
Congestive Heart Failure, n (%)	2768 (45.48%)	243 (70.43%)	0.00000
Dementia, n (%)	398 (6.54%)	49 (14.20%)	0.00000
Diabetes without Complications, n (%)	1313 (21.57%)	77 (22.32%)	0.79510
Paraplegia, n (%)	316 (5.19%)	54 (15.65%)	0.00000
Metastatic Solid Tumor, n (%)	192 (3.15%)	39 (11.30%)	0.00000
Readmission within 60 days, n (%)	1249 (20.52%)	129 (34.78%)	0.00000
Marital Status: Widowed, n (%)	1051 (17.27%)	74 (21.45%)	0.05546
Observation Admission, n (%)	2702 (44.40%)	103 (29.86%)	0.00000
Surgical Admission, n (%)	418 (6.87%)	5 (1.45%)	0.00012
Microbiology: Negative, n (%)	5113 (84.01%)	191 (55.36%)	0.00000
No applied procedures, n (%)	2197 (36.10%)	53 (15.36%)	0.00000
Heparin, n (%)	3984 (65.46%)	313 (90.72%)	0.00000
Obesity, n (%)	831 (13.65%)	37 (10.72%)	0.14204
Anion Gap [mEq/L], mean $\pm$ SD	15.89 $\pm$ 3.55	18.33 $\pm$ 4.27	0.00000
Bicarbonate [mEq/L], mean $\pm$ SD	27.79 $\pm$ 4.04	30.58 $\pm$ 5.30	0.00000
Creatinine [mg/dL], mean $\pm$ SD	1.43 $\pm$ 1.28	2.05 $\pm$ 1.87	0.00000
Hemoglobin [g/dL], mean $\pm$ SD	12.45 $\pm$ 2.07	11.6 $\pm$ 2.08	0.00000

Table 1. Cont.

Feature	Survivors (n = 6086)	Non-Survivors (n = 345)	p-Value (Two-Sided)
Potassium [mEq/L], mean $\pm$ SD	4.69 $\pm$ 0.71	5.26 $\pm$ 0.98	0.00000
Sodium [mEq/L], mean $\pm$ SD	141.88 $\pm$ 4.01	144.94 $\pm$ 5.67	0.00000
Urea Nitrogen [mg/dL], mean $\pm$ SD	30.76 $\pm$ 21.11	48.97 $\pm$ 27.20	0.00000
White Blood Cells [$10^6/\mu\text{L}$], mean $\pm$ SD	12.3 $\pm$ 9.55	17.35 $\pm$ 10.19	0.00000
Sum of ICU Length of Stay [days], median (Q1,Q3)	0 (0, 1.45)	3.19 (1.68, 7.73)	0.00000
Number of Diagnoses, median (Q1,Q3)	17 (12,22)	27 (21,33)	0.00000
Charlson Comorbidity Index, median (Q1,Q3)	5 (4,7)	8 (6,10)	0.00000
Hospital Days [days], median (Q1,Q3)	5 (2,9)	13 (8,23)	0.00000

3.2. Machine Learning

ROC–AUC curves were presented for all five algorithms used during training (see Figure 1). As initially observed, the results obtained by LightGBM, XGBoost, and Random Forest are very similar on the ROC curves, suggesting comparable effectiveness of these algorithms in the analyzed task. The primary factors determining the final algorithm selection were MCC and AUC, as these metrics effectively evaluate the quality of the classification model in the context of a highly imbalanced dataset. Table 2 summarizes all metrics for each model, along with the 95% confidence intervals (CI) for the results obtained during internal cross-validation.

Table 2. Performance metrics on training dataset. Results of cross-validation for five different machine learning algorithms.

Algorithm	AUC (95%CI)	Accuracy (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	MCC (95%CI)
LightGBM	0.873 (0.865–0.881)	0.862 (0.856–0.867)	0.886 (0.870–0.901)	0.860 (0.854–0.866)	0.439 (0.427–0.450)
XGBoost	0.871 (0.863–0.879)	0.846 (0.840–0.853)	0.899 (0.882–0.915)	0.843 (0.836–0.850)	0.422 (0.411–0.433)
RF	0.862 (0.854–0.870)	0.853 (0.846–0.859)	0.873 (0.855–0.890)	0.851 (0.845–0.858)	0.419 (0.409–0.429)
Adaboost	0.844 (0.839–0.849)	0.764 (0.745–0.783)	0.934 (0.908–0.961)	0.754 (0.733–0.775)	0.348 (0.340–0.357)
LR	0.795 (0.785–0.805)	0.770 (0.765–0.776)	0.822 (0.801–0.842)	0.767 (0.762–0.773)	0.301 (0.291–0.312)

Table 3 presents the prediction results on the test set obtained by the two best-performing models during training. The similar results achieved by the closely related algorithms, LightGBM and XGBoost, are not surprising; both algorithms are based on gradient boosting and share many underlying principles and techniques [22,23]. However, the results obtained by LightGBM were superior. All metrics showed higher values, with particular emphasis on the model's high sensitivity. Given the importance of predicting patient mortality, maintaining high sensitivity is crucial, while also achieving satisfactory specificity results. The Z-test performed on the results obtained by the LightGBM and XG-

Boost models on the test data did not reveal statistically significant differences in sensitivity or specificity ($p > 0.05$). Compared to the CHA₂DS₂-VASc score (mean 3.01, SD 1.318), which achieved an AUC of 0.639 (95% CI: 0.611–0.665) for mortality prediction, the results obtained by the presented models were substantially superior [24,25].

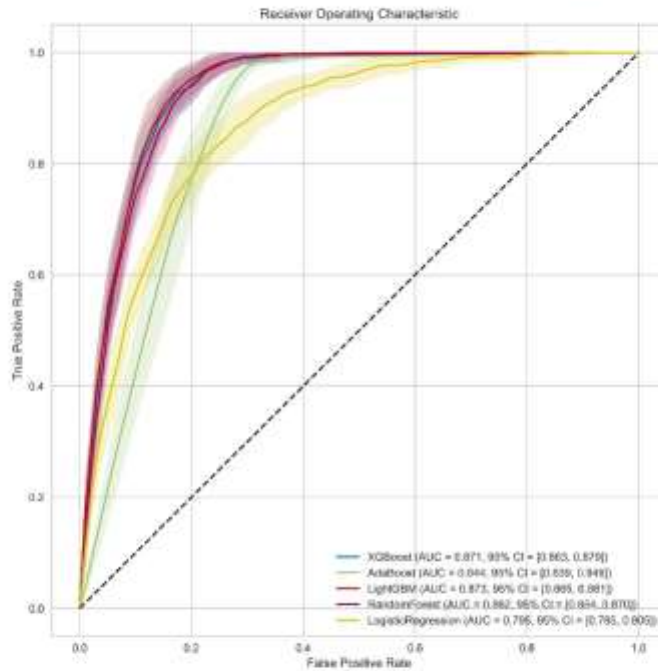


Figure 1. ROC curves visualizing performance of five different machine learning algorithms.

Table 3. Performance metrics on test dataset. Predictions made by LightGBM and XGBoost.

Algorithm	AUC (95%CI)	Accuracy (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	MCC (95%CI)
LightGBM	0.886 (0.849–0.919)	0.862 (0.842–0.880)	0.913 (0.842–0.973)	0.859 (0.839–0.878)	0.450 (0.388–0.507)
XGBoost	0.873 (0.834–0.908)	0.849 (0.828–0.868)	0.899 (0.823–0.967)	0.846 (0.825–0.866)	0.425 (0.362–0.483)

3.3. Shapley Values

To explain the obtained results, the SHAP library was used, which employs Shapley values to identify the features with the greatest impact on predictions both globally and locally [26,27]. A beeswarm plot, as presented in Figure 2, visualizes the impact of each feature on the model's predictions using SHAP values. Each point on the plot represents

a SHAP value for a feature and an individual prediction. The position on the x-axis indicates the magnitude and direction of the feature's impact on the prediction, while the color represents the feature value. This plot provides a comprehensive overview of which features are most influential in the model's decision-making process and how they contribute to the predictions. As shown, the most important features include the length of stay in the ICU and the hospital, as well as comorbidities. Besides laboratory test results, these factors directly inform us about the clinical condition of patients with AF. Additional features that improved the model's performance include the type of admission, the absence of performed procedures, and negative microbiology test results. Age and gender also have a significant impact on the model's predictions. The only medication among the selected 27 features is heparin treatment, which inhibits the blood coagulation process primarily by activating antithrombin (AT), a potent inhibitor of thrombin and a weaker inhibitor of other plasma coagulation factors (Xa, IXa, XIa, XIIa).

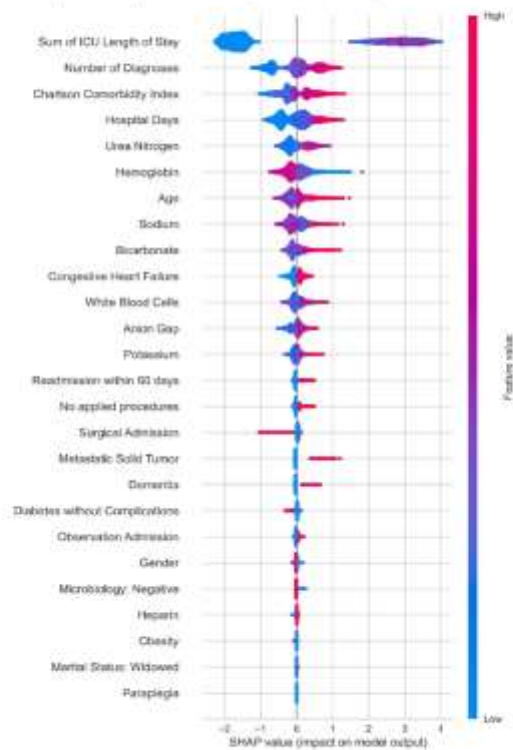


Figure 2. Beeswarm Plot: This plot visualizes the distribution of SHAP values across all features in the dataset, highlighting the impact of each feature on the model's output.

4. Discussion

There is a lack of machine learning models capable of predicting mortality in patients using DOACs for AF. According to ESC recommendations, these drugs are the first choice for preventing thromboembolic events in this patient group [28]. The use of DOACs in the treatment of patients with AF is not only associated with significant therapeutic benefits but has also been shown to be safe in numerous studies, particularly with respect to the risk of major bleeding and intracranial hemorrhage. For instance, the XANTUS study—a prospective, real-world registry including over 6700 patients with AF treated with rivaroxaban—reported a major bleeding rate of 2.1 per 100 patient-years and confirmed a low incidence of fatal bleeding events [29]. Similarly, recent data from an Italian multicenter registry demonstrated that DOACs, including rivaroxaban, were safe and effective in routine clinical practice, with no major bleeding complications observed during a 12-month follow-up period [30]. These findings are further corroborated by large-scale analyses from Danish cohorts, which have shown that DOACs are associated with a lower risk of death and major bleeding compared to warfarin, supporting their favorable safety profile in real-world settings [31]. Since the MIMIC-IV database also contains records from before 2010, a period before the widespread use of DOACs in AF patients, the number of patients in this group is relatively low. A significant challenge is the high imbalance of the dataset, which necessitates the use of algorithms capable of handling such data and the selection of specific quality metrics. Mortality studies based on this database often focus on patients in the ICU, primarily due to the abundance of data for these patients [32–34]. In the context of the MIMIC database, ICU is considered as a stay in any intensive care unit, without distinguishing whether it was an MICU (Medical Intensive Care Unit), an SICU (Surgical Intensive Care Unit), a CCU (Cardiac Care Unit), a CSRU (Cardiac Surgery Recovery Unit), or an NICU (Neonatal Intensive Care Unit). All these units are collectively referred to as ICU in the MIMIC database [19,35]. Developing such a model enables the implementation of appropriate clinical measures, such as closer monitoring of the patient's condition to reduce the risk of death. The model is not limited to predicting death from a specific cause due to the limited scope of the data. An additional limitation is the availability of data from only one facility. Following the example of the MIMIC database, it would be beneficial for other large institutions to develop similar datasets. This would reduce data bias and improve the testing of developed models. Another important limitation is the use of a simplified preprocessing approach for handling missing data by removing incomplete records. While this reduces complexity and maintains a streamlined workflow, it introduces the risk of bias. Exploring the impact of various data imputation techniques would be an interesting direction for future research. Preliminary attempts to apply the Multiple Imputation by Chained Equations (MICE) method did not yield a significant improvement in results. Additionally, the study does not employ more complex algorithms or ensemble methods, which could potentially improve performance but often at the expense of model interpretability. The significant risk factors for mortality within six months identified by the model for this patient group can serve as a basis for further research, including clinical studies. In the future, this could lead to improved personalized treatment for AF patients. It is crucial to interpret the results correctly; the model indicating heparin does not imply that heparin increases the risk of death. Instead, it should be interpreted as a reflection of the clinical condition of the patient requiring such treatment according to current guidelines. It is important to remember the crucial role that proper adherence plays in the use of DOACs, which is influenced by various factors, including the patient's age and social status. Effective communication between patients and healthcare providers is crucial for improving treatment adherence. This involves clearly explaining the disease and the necessity of the treatment, as well as understanding the patient's lifestyle, physical

and financial capabilities, and preferences. Involving the patient in the decision-making process regarding anticoagulant therapy leads to a better understanding and acceptance of the therapy goals [36]. The selected features serve as descriptors of the patient's clinical status and should not be interpreted as independent markers equivalent to those obtained from clinical studies. Artificial intelligence models should be utilized by specialists who are capable of interpreting the presented results and using them as a supportive tool for assessing the clinical condition of a patient, merely highlighting important factors that may increase or decrease the risk of mortality in a given individual. Model indications, such as the identification of heparin use as significant, should not be regarded as recommendations for treatment modification. Instead, such findings should be interpreted as reflecting the clinical status of the patient, which may require specific therapeutic interventions based on current medical knowledge and established clinical guidelines. False-positive predictions may lead to unnecessary interventions, such as increased monitoring, additional medications, or invasive procedures. These actions could expose patients to avoidable side effects, increased healthcare costs, and psychological distress due to being labeled as high-risk. Overestimating mortality risk could divert limited healthcare resources (e.g., ICU beds or specialist attention) away from patients who genuinely need them, potentially compromising care for others. On the other hand, false-negative predictions may result in high-risk patients being overlooked for necessary interventions, leading to preventable adverse outcomes, including death. This undermines the primary goal of the predictive model. If patients or clinicians perceive the model as unreliable, it could erode trust in AI-based tools, hindering their adoption in clinical practice. Ethical considerations demand a careful balance between sensitivity (minimizing false negatives) and specificity (minimizing false positives). For mortality prediction, prioritizing sensitivity may be more ethical to ensure high-risk patients are not missed, but this must be weighed against the risks of overtreatment. Predictions should be used to support, not replace, clinical judgment. Patients must be informed about the limitations of the model, including the possibility of false predictions, to ensure they can make autonomous decisions about their care. False predictions may disproportionately affect certain patient groups if the model is biased due to imbalanced training data. This raises ethical concerns about fairness and equity in healthcare delivery. One of the features used by the model is obesity, which has been studied for its impact on the achieved concentrations of DOACs. This can result in either a lack of therapeutic effect or an increased risk of bleeding [37,38]. Another study focused on the direct impact of obesity on short- and medium-term mortality in critically ill patients with AF, demonstrating a lack of linear relationship between BMI and all-cause mortality [39].

In medical data analysis for predicting patient mortality, the application of various machine learning algorithms offers diverse benefits. Logistic Regression is a simple and interpretable model that allows for quick insights into feature significance and performs well with linearly separable data, serving as a solid baseline. AdaBoost, as a boosting method, effectively identifies difficult cases and improves model accuracy while reducing the risk of overfitting by limiting the depth of base classifiers. XGBoost, on the other hand, is known for its high performance and accuracy due to gradient optimization and built-in regularization, making it a popular choice for predictive tasks, especially with large datasets. LightGBM, an alternative to XGBoost, provides faster training and more efficient memory usage, which is crucial with a large number of variables and class imbalance. Random Forest offers stability and noise resistance by aggregating the results of multiple decision trees and allows for feature importance evaluation, supporting result interpretation in a clinical context. The selected set of algorithms is well established in numerous scientific studies utilizing binary classification to predict patient mortality across various conditions [40–44].

The use of Explainable AI (XAI) is essential for such solutions in medicine. Without the ability to trace the model's reasoning, it is challenging to gain the trust of healthcare providers. SHAP values are effective in this context, both globally and locally. Local explanations for individual patients, indicating what drives the model's predictions, not only enhance the trust of medical professionals but also enable personalized approaches. See Figure 3.

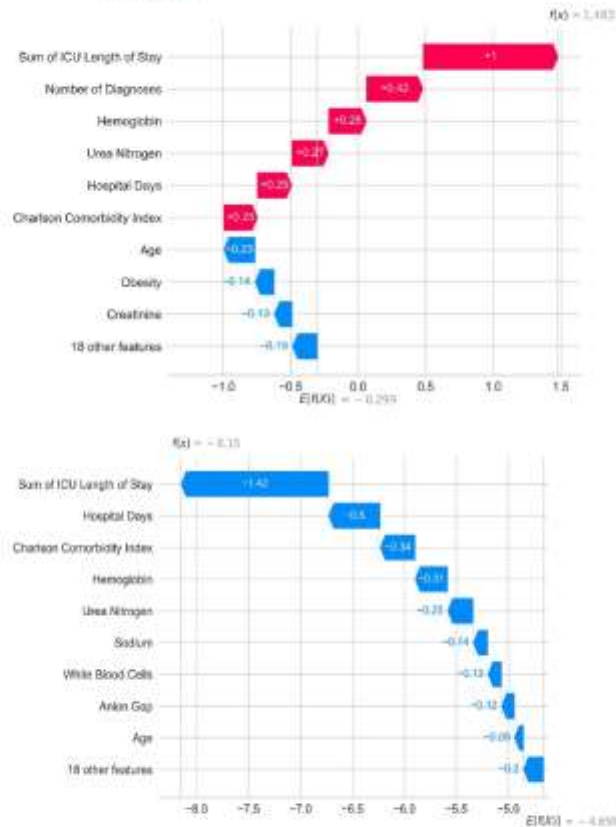


Figure 3. Waterfall Plot: This plot provides a detailed breakdown of SHAP values for two individual patients with different prediction outcomes. The first patient was classified by the model as high risk for mortality, while the second patient was predicted to be at low risk.

To the best of our knowledge, this is the first project aimed at predicting mortality in this specific patient group, and therefore we cannot compare our results with other studies. The MIMIC database itself is widely used for predicting patient mortality, including those

with AF [45]. However, the target population in our study is different, making direct comparison of the obtained results impossible.

Although the model achieves highly satisfactory results, clinical judgment remains essential in interpreting and applying its predictions. It is important to remember that such models are tools to support daily work in clinical settings.

5. Conclusions

This study demonstrates the potential of using machine learning to predict mortality in AF patients using DOACs within six months of hospital discharge. The best results were achieved by the model based on the LightGBM algorithm. Utilizing AI-based tools can save clinical time and facilitate personalized disease management. However, further research is needed, preferably multi-center studies, to improve the model and validate it on entirely external datasets.

Author Contributions: Conceptualization, L.L., M.L. and G.G.; methodology, L.L.; validation, L.L.; data curation, L.L. and M.L.; writing—original draft preparation, L.L. and M.L.; writing—review and editing, E.G. and G.G.; visualization, L.L. and M.L.; supervision, E.G. and G.G.; project administration, G.G.; funding acquisition, E.G. and G.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study's retrospective data.

Informed Consent Statement: Patient consent was waived due to data from an anonymized database.

Data Availability Statement: Data is available on reasonable request.

Acknowledgments: The authors would like to thank Szymon Łodziński for his technical assistance in database management.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AdaBoost	Adaptive Boosting
AF	Atrial fibrillation
AI	Artificial Intelligence
AT	Antithrombin
AUC	Area Under the Curve
BMI	Body Mass Index
CCI	Charlson Comorbidity Index
CCU	Cardiac Care Unit
CI	Confidence Interval
CITI	Collaborative Institutional Training Initiative
CSRU	Cardiac Surgery Recovery Unit
DOACs	Direct Oral Anticoagulants
DUA	Data Use Agreement
ICD	International Classification of Diseases
ICU	Intensive Care Unit
KNN	k-Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
LightGBM	Light Gradient-Boosting Machine
LR	Logistic Regression

MCC	Matthews Correlation Coefficient
MICE	Multiple Imputation by Chained Equations
MICU	Medical Intensive Care Unit
MIMIC	Medical Information Mart for Intensive Care
NICU	Neonatal Intensive Care Unit
RF	Random Forest
ROC	Receiver Operating Characteristic
SD	Standard deviation
SHAP	Shapley Additive exPlanations
SICU	Surgical Intensive Care Unit
VKAs	traditional vitamin K antagonists
XAI	Explainable Artificial Intelligence
XGBoost	eXtreme Gradient Boosting

References

1. Massera, D.; Wang, D.; Vorchheimer, D.A.; Negassa, A.; Garcia, M.J. Increased Risk of Stroke and Mortality Following New-Onset Atrial Fibrillation during Hospitalization. *Emergancz* **2017**, *19*, 929–936. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Patel, P.J.; Katz, R.; Borovsky, Y.; Killian, A.; Levine, J.M.; McNaughton, N.W.; Collins, D.; Supple, G.; Dixit, S.; Epstein, A.E.; et al. Race and Stroke in an Atrial Fibrillation Inception Cohort: Findings from the Penn Atrial Fibrillation Free Study. *Heart Rhythm* **2018**, *15*, 487–493. [\[CrossRef\]](#)
3. Thrall, G.; Lane, D.; Carroll, D.; Lip, G.Y.H. Quality of Life in Patients with Atrial Fibrillation: A Systematic Review. *Am. J. Med.* **2006**, *119*, 448.E1–448.E19. [\[CrossRef\]](#)
4. Go, A.S.; Hylek, E.M.; Phillips, K.A.; Chang, Y.; Henault, L.E.; Selby, J.V.; Singer, D.E. Prevalence of Diagnosed Atrial Fibrillation in Adults. *JAMA* **2001**, *285*, 2370. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Ball, J.; Carrington, M.J.; McMurray, J.J.V.; Stewart, S. Atrial Fibrillation: Profile and Burden of an Evolving Epidemic in the 21st Century. *Int. J. Cardiol.* **2013**, *167*, 1807–1824. [\[CrossRef\]](#)
6. Kaloniz, Z.; Sredniawa, B.; Mitrega, K.; Wierucki, L.; Sokal, A.; Lip, G.; Bendosz, P.; Stokwiszewski, J.; Boldol, J.; Zieloniewicz, P.; et al. Prevalence of Atrial Fibrillation in the 65 or over Polish Population. Report of Cross-Sectional NOMED-AF Study. *Kardiol. Pol.* **2023**, *81*, 14–21. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Steffel, J.; Collins, R.; Arzi, M.; Cornu, T.; Desteghe, L.; Haensler, K.G.; Olsson, J.; Reinecke, H.; Roldan-Schilling, V.; Rowell, N.; et al. 2021 European Heart Rhythm Association Practical Guide on the Use of Non-Vitamin K Antagonist Oral Anticoagulants in Patients with Atrial Fibrillation. *EP Eur.* **2021**, *23*, 1612–1676. [\[CrossRef\]](#)
8. Grzesik, G.; Rogowicz, D.; Wolowicz, L.; Ratajczak, A.; Galewski, W.; Chudzińska, M.; Sinkiewicz, A.; Banach, J. The Clinical Significance of Drug–Food Interactions of Direct Oral Anticoagulants. *Int. J. Mol. Sci.* **2021**, *22*, 8531. [\[CrossRef\]](#)
9. Grzesik, G. Therapeutic Monitoring of Direct Oral Anticoagulants—an 8-Year Observational Study. *Acta Haematol. Pol.* **2023**, *52*, 446–452. [\[CrossRef\]](#)
10. Ledziński, L.; Grzesik, G. Artificial Intelligence Technologies in Cardiology. *J. Cardiovasc. Dev. Dis.* **2023**, *10*, 202. [\[CrossRef\]](#)
11. Xu, Q.; Peng, Y.; Tan, J.; Zhao, W.; Yang, M.; Tan, J. Prediction of Atrial Fibrillation in Hospitalized Elderly Patients With Coronary Heart Disease and Type 2 Diabetes Mellitus Using Machine Learning: A Multicenter Retrospective Study. *Front. Public Health* **2022**, *10*, 842104. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Hu, W.; Lin, C. Real-World Observational Study of Assessment of CHA2DS2-VASc, C2HEST and HAVOC Scores for Atrial Fibrillation among Patients with Rheumatological Disorders: A Nationwide Analysis. *Postgrad. Med. J.* **2022**, *98*, 837–841. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Fox, K.A.A.; Vardoulakis, S.; Pogue, K.S.; Bassand, J.-P.; Camm, A.J.; Fitzmaurice, D.A.; Goldhaber, S.Z.; Goto, S.; Haas, S.; Kayani, G.; et al. GARFIELD-AF Risk Score for Mortality, Stroke, and Bleeding within 2 Years in Patients with Atrial Fibrillation. *Eur. Heart J. Qual. Care Clin. Outcomes* **2022**, *8*, 214–227. [\[CrossRef\]](#)
14. Hu, W.; Lin, C. Comparisons of HATCH, HAVOC and CHA2DS2-VASc Scores for All-Cause Mortality Prediction in Atrial Fibrillation: A Real-World Evidence Study. *Postgrad. Med. J.* **2023**, *99*, 326–332. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Wang, B.; Jin, F.; Cao, H.; Li, Q.; Zhang, P. Development of a Risk Score for Predicting One-Year Mortality in Patients with Atrial Fibrillation Using XGBoost-Assisted Feature Selection. *Pol. Heart J.* **2024**, *62*, 941–948. [\[CrossRef\]](#)
16. Liu, X.; Jiang, J.; Wei, L.; Xing, W.; Shang, H.; Liu, G.; Liu, B. Prediction of All-Cause Mortality in Coronary Artery Disease Patients with Atrial Fibrillation Based on Machine Learning Models. *BMC Cardiovasc. Disord.* **2021**, *21*, 499. [\[CrossRef\]](#)

17. Falsetti, L.; Rucen, M.; Prioretti, M.; Viticchi, G.; Zaccaro, V.; Scarpioni, M.; Giannelli, L.; Meconzi, G.; Nitti, C.; Salvi, A. Risk Prediction of Clinical Adverse Outcomes with Machine Learning in a Cohort of Critically Ill Patients with Atrial Fibrillation. *Sci. Rep.* **2021**, *11*, 18925. [\[CrossRef\]](#)
18. Alistair, J.; Lucas, B.; Tom, P.; Brian, C.; Benjamin, M.; Steven, H.; Lee, A.C.; Roger, M. *MIMIC-IV, version 3.0*; PhysioNet: Boston, MA, USA, 2024.
19. Johnson, A.E.W.; Bulgarelli, L.; Shen, L.; Gayles, A.; Shammout, A.; Hoernig, S.; Pollard, T.J.; Hao, S.; Moody, B.; Gow, B.; et al. MIMIC-IV, a Freely Accessible Electronic Health Record Dataset. *Sci. Data* **2023**, *10*, 1. [\[CrossRef\]](#)
20. Goodberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* **2000**, *102*, E213–E220. [\[CrossRef\]](#)
21. Bergstra, J.; Yamins, D.; Cox, D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; PMLR: Atlanta, GA, USA, 2013; Volume 28, pp. 115–123.
22. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.
23. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3146–3154.
24. Xia, Y.; Liang, A.; Wang, M.; Zhang, J. Risk Analysis of the Association between EASIX and All-Cause Mortality in Critical Ill Patients with Atrial Fibrillation: A Retrospective Study from MIMIC-IV Database. *Eur. J. Med. Res.* **2025**, *30*, 344. [\[CrossRef\]](#)
25. Morone, D.; Kroep, S.; Ricci, F.; Renda, G.; Patti, G.; Kirchhof, P.; Chuang, L.-H.; van Hoot, B.; De Caterina, R. Mortality Prediction of the CHA2DS2-VASc Score, the HAS-BLED Score, and Their Combination in Anticoagulated Patients with Atrial Fibrillation. *J. Clin. Med.* **2020**, *9*, 3987. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3146–3154.
27. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Van Gelder, I.C.; Rienstra, M.; Bunting, K.V.; Casado-Arroyo, R.; Caso, V.; Crijs, H.J.G.M.; De Potter, T.J.R.; Dwight, J.; Guasti, L.; Hanke, T.; et al. 2024 ESC Guidelines for the Management of Atrial Fibrillation Developed in Collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *Eur. Heart J.* **2024**, *45*, 3314–3414. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Camm, A.J.; Amaretti, P.; Hiss, S.; Hess, S.; Kirchhof, P.; Kuhl, S.; van Eickels, M.; Turpie, A.G.G. XANTUS: A Real-World, Prospective, Observational Study of Patients Treated with Rivaroxaban for Stroke Prevention in Atrial Fibrillation. *Eur. Heart J.* **2016**, *37*, 1145–1153. [\[CrossRef\]](#)
30. Lavalley, C.; Pierucci, N.; Mariani, M.V.; Firo, A.; Borrelli, A.; Grimaldi, M.; Rossillo, A.; Notarstefano, P.; Compagnucci, P.; Russo, A.D.; et al. Italian Registry in the Setting of Atrial Fibrillation Ablation with Rivaroxaban-IRIS. *Minerva Cardiol. Angiol.* **2024**, *72*, 625–637. [\[CrossRef\]](#)
31. Larsen, T.B.; Skjott, F.; Nielsen, P.B.; Kjeldgaard, J.N.; Lip, G.Y.H. Comparative Effectiveness and Safety of Non-Vitamin K Antagonist Oral Anticoagulants and Warfarin in Patients with Atrial Fibrillation: Propensity Weighted Nationwide Cohort Study. *BMJ* **2016**, *353*, i3189. [\[CrossRef\]](#)
32. Luo, Y.; Dong, R.; Liu, J.; Wu, B. A Machine Learning-Based Predictive Model for the in-Hospital Mortality of Critically Ill Patients with Atrial Fibrillation. *Int. J. Med. Inform.* **2024**, *191*, 105585. [\[CrossRef\]](#)
33. Li, Q.; Nie, J.; Cao, M.; Luo, C.; Sun, C. Association between Inflammation Markers and All-Cause Mortality in Critical Ill Patients with Atrial Fibrillation: Analysis of the Multi-Parameter Intelligent Monitoring in Intensive Care (MIMIC-IV) Database. *J. Heart Vasc.* **2024**, *51*, 101372. [\[CrossRef\]](#)
34. Hu, Y.; Zhao, Y.; Zhang, J.; Li, C. The Association between Triglyceride Glucose-Body Mass Index and All-Cause Mortality in Critically Ill Patients with Atrial Fibrillation: A Retrospective Study from MIMIC-IV Database. *Cardiovasc. Diabetol.* **2024**, *23*, 64. [\[CrossRef\]](#)
35. Johnson, A.E.W.; Pollard, T.J.; Shen, L.; Lehman, L.H.; Feng, M.; Ghaemi, M.; Moody, B.; Szolovits, P.; Anthony Cell, L.; Mark, R.G. MIMIC-III, a Freely Accessible Critical Care Database. *Sci. Data* **2016**, *3*, 16005. [\[CrossRef\]](#)
36. Grzesk, C.; Janiszewska, E.; Malinowski, B.; Kubica, A.; Wiciński, M. Adherence in Patients with Atrial Fibrillation Treated with Dabigatran. *Kardiol. Pol.* **2018**, *76*, 1562–1563. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Wołowicz, L.; Kusiak, M.; Budzyński, J.; Wołowicz, A.; Jasiński, A.; Wiciński, M.; Pedrycz-Wieczorska, A.; Rogowicz, D.; Grzesk, G. Therapeutic Drug Monitoring of Direct Oral Anticoagulants in Patients with Extremely Low and High Body Weight—Pilot Study. *J. Clin. Med.* **2023**, *12*, 4969. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Whittermore, H.; Posen, A.K.; Hollenbart, F.E.; Groo, V.; Wenzler, E.; Tihon, J.J. The Impact of Body Weight and Renal Function on the Risk of Bleeding with Direct Oral Anticoagulants in Atrial Fibrillation. *Ann. Pharmacother.* **2021**, *55*, 1309–1317. [\[CrossRef\]](#)

39. Yang, D.; Ye, S.; Zhang, K.; Huang, Z.; Zhang, L. Association between Obesity and Short- and Medium-Term Mortality in Critically Ill Patients with Atrial Fibrillation: A Retrospective Cohort Study. *BMC Cardiovasc. Disord.* **2023**, *23*, 150. [\[CrossRef\]](#)
40. Son, B.; Myung, J.; Shin, Y.; Kim, S.; Kim, S.H.; Chung, J.-M.; Noh, J.; Cho, J.; Chung, H.S. Improved Patient Mortality Predictions in Emergency Departments with Deep Learning Data-Synthesis and Ensemble Models. *Sci. Rep.* **2023**, *13*, 15031. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Hu, T.; Chan, C.; Wu, C.; Chien, T.; Li, C. Machine Learning-Based Predictions of Mortality and Readmission in Type 2 Diabetes Patients in the ICU. *Appl. Sci.* **2024**, *14*, 8443. [\[CrossRef\]](#)
42. Li, X.; Wu, R.; Zhao, W.; Shi, R.; Zhu, Y.; Wang, Z.; Pan, H.; Wang, D. Machine Learning Algorithm to Predict Mortality in Critically Ill Patients with Sepsis-Associated Acute Kidney Injury. *Sci. Rep.* **2023**, *13*, 5223. [\[CrossRef\]](#)
43. Sharifi-Kia, A.; Nahvijou, A.; Shekhtaheri, A. Machine Learning-Based Mortality Prediction Models for Smoker COVID-19 Patients. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 129. [\[CrossRef\]](#)
44. Gao, J.; Lu, Y.; Ashrafi, N.; Dumlugu, I.; Alael, K.; Pishgar, M. Prediction of Sepsis Mortality in ICU Patients Using Machine Learning Methods. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 228. [\[CrossRef\]](#)
45. Wang, Z.; Zhang, L.; Chao, Y.; Xu, M.; Gong, X.; Hu, X. DEVELOPMENT OF A MACHINE LEARNING MODEL FOR PREDICTING 28-DAY MORTALITY OF SEPTIC PATIENTS WITH ATRIAL FIBRILLATION. *Shock* **2023**, *59*, 400–406. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

10.3. Publikacja III

Article

Personalized Medicine in Pulmonary Arterial Hypertension: Utilizing Artificial Intelligence for Death Prevention

Lukasz Ledziński ¹, Grzegorz Grzesik ^{1,*}, Michał Ziolkowski ¹, Marcin Waligóra ^{2,3,4}, Marcin Kurzyński ⁵, Tatiana Mularek-Kubzdela ⁶, Anna Smukowska-Gorynia ⁶, Ilona Skoczylas ⁷, Łukasz Chrzanowski ⁸, Piotr Błaszczak ⁹, Miłosz Jaguszewski ¹⁰, Beata Kuśmierczyk-Drożdż ¹¹, Katarzyna Ptaszyńska ¹², Katarzyna Mizia-Steć ¹³, Ewa Malinowska ¹⁴, Małgorzata Peregud-Pogorzelska ¹⁵, Ewa Lewicka ¹⁶, Michał Tomaszewski ¹⁷, Wojciech Jachet ¹⁸, Michał Florczyk ¹⁹, Ewa Mroczek ¹⁹, Zbigniew Gąsior ²⁰, Agnieszka Pawlak ^{21,22}, Katarzyna Betkier-Lipińska ²³, Piotr Prusaczyk ²⁴, Olga Dzikowska-Diduch ²⁴, Katarzyna Widejko ²⁵, Judyta Winowska-Józwa ²⁶ and Grzegorz Kopeć ^{2,3}

¹ Department of Cardiology and Clinical Pharmacology, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, 85-168 Bydgoszcz, Poland; 503345@kolosari.pl (L.L.); boldynski@wp.pl (M.Z.)

² Department of Cardiac and Vascular Diseases, St. John Paul II Hospital in Kraków, 31-202 Kraków, Poland

³ Pulmonary Circulation Centre, Department of Cardiac and Vascular Diseases, Faculty of Medicine, Jagiellonian University Medical College, 31-202 Kraków, Poland

⁴ Center for Innovative Medical Education, Department of Medical Education, Faculty of Medicine, Jagiellonian University Medical College, 31-202 Kraków, Poland

⁵ Department of Pulmonary Circulation, Thromboembolic Diseases and Cardiology, Centre of Postgraduate Medical Education, Fryderyk Chopin Hospital in Turpin Health Centre Otwock, 05-400 Otwock, Poland; marcin.kurzynski@poczta.onet.pl (M.K.); michal.florczyk@poczta.onet.pl (M.F.)

⁶ Department of Cardiology, Poznań University of Medical Sciences, 61-701 Poznań, Poland

⁷ 2nd Department of Cardiology, Faculty of Medical Sciences in Zabrze, Medical University of Silesia, 41-800 Katowice, Poland

⁸ Cardiology Department, Medical University of Łódź, 91-347 Łódź, Poland; chrzanowski@poczta.onet.pl

⁹ Department of Cardiology, Cardinal Wyszyński Hospital, 20-718 Lublin, Poland

¹⁰ 1st Department of Cardiology, Medical University of Gdańsk, 80-210 Gdańsk, Poland

¹¹ Department of Congenital Heart Disease, Institute of Cardiology, 04-628 Warsaw, Poland

¹² Department of Cardiology, Medical University of Białystok, 15-276 Białystok, Poland

¹³ Centre of the European Reference Network for Rare, Low Prevalence or Complex Diseases of the Heart (ERN GUARD Heart), First Department of Cardiology, School of Medicine in Katowice, Medical University of Silesia in Katowice, 40-635 Katowice, Poland

¹⁴ Pulmonary Department, University of Warmia and Mazury, 10-337 Olsztyn, Poland

¹⁵ Department of Cardiology, Pomeranian Medical University in Szczecin, 70-111 Szczecin, Poland; malgorzata.peregud-pogorzelska@pum.edu.pl

¹⁶ Department of Cardiology and Electrophysiology, Medical University of Gdańsk, 80-211 Gdańsk, Poland; ewa@pum.edu.pl

¹⁷ Department of Cardiology, Medical University of Lublin, 20-080 Lublin, Poland

¹⁸ 2nd Department of Cardiology, Faculty of Medical Sciences in Zabrze, Medical University of Silesia in Katowice, 41-800 Zabrze, Poland

¹⁹ Clinic of Heart Diseases, Institute of Heart Diseases, University Clinical Hospital in Wrocław, 50-296 Wrocław, Poland

²⁰ Department of Cardiology, School of Health Sciences in Katowice, Medical University of Silesia in Katowice, 40-635 Katowice, Poland

²¹ Department of Cardiology, National Medical Institute of the Ministry of the Interior and Administration, 02-507 Warszawa, Poland; a.pawlak@wp.pl

²² Clinical and Research Department of Applied Physiology, Mossakowski Medical Research Institute, Polish Academy of Sciences, 02-106 Warszawa, Poland

²³ Department of Cardiology and Internal Medicine, Military Institute of Medicine—National Research Institute, 04-111 Warszawa, Poland; kbetkier-lipinska@um.mil.pl

²⁴ Department of Internal Medicine and Cardiology with the Center for Diagnosis and Treatment of Venous Thromboembolism, Medical University of Warsaw, 02-003 Warszawa, Poland

²⁵ Department of Cardiology, Copper Health Center, 29-300 Lubin, Poland

²⁶ Department of Cardiology, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, 85-168 Bydgoszcz, Poland



Academic Editor: Matthias Uebachs

Received: 18 October 2023

Revised: 15 November 2023

Accepted: 21 November 2023

Published: 25 November 2023

Citation: Ledziński, L.; Grzesik, G.; Ziolkowski, M.; Waligóra, M.; Kurzyński, M.; Mularek-Kubzdela, T.; Smukowska-Gorynia, A.; Skoczylas, I.; Chrzanowski, Ł.; Błaszczak, P. et al. Personalized Medicine in Pulmonary Arterial Hypertension: Utilizing Artificial Intelligence for Death Prevention. *J. Clin. Med.* **2023**, *14*, 8125. <https://doi.org/10.3390/jcm14238125>

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

[‡] Department of Cardiology, Provincial Specialist Hospital in Szczecin, 70-111 Szczecin, Poland; jowianski@poczta.onet.pl

* Correspondence: g.grodek@umk.pl

Abstract

Background/Objectives: Pulmonary arterial hypertension (PAH) is a complex cardiovascular disease with a high burden of morbidity and mortality. Although several risk prediction models have been proposed, the exact significance of distinct clinical parameters in predicting survival in PAH remains unclear. It is important to emphasize that this study does not aim to validate or contradict existing clinical risk assessment calculators provided by the ESC or other scientific societies. Instead, the goal of this research is to identify and rank clinical parameters according to their importance in predicting mortality in PAH patients using machine learning techniques. **Methods:** Using the Database of Pulmonary Hypertension in the Polish population (BNP-PL) registry, 1755 adult patients with PAH were selected. Feature engineering was conducted using domain knowledge, guided by European Society of Cardiology (ESC) recommendations. Features were reduced using LASSO regression and sequential feature elimination algorithms. A classification model was built using the XGBoost algorithm, utilizing 17 features. The model was tested on a preselected subset of the BNP-PL data. The Shapley Additive Explanations (SHAP) method was used to explain the model's predictions and to rank feature importance. **Results:** The model achieved satisfactory results across evaluated metrics, including an area under the curve of 0.767, accuracy of 0.738, specificity of 0.733, and sensitivity of 0.800. SHAP values effectively ranked the features, corroborating the significance of parameters present in the ESC risk stratification tables. Furthermore, local interpretation of results using SHAP enabled individualized assessment of feature importance, enhancing clinical applicability. **Conclusions:** The proposed artificial intelligence-based model demonstrates satisfactory predictive capability, highlighting the potential of machine learning techniques to support more personalized approaches to the management of PAH patients. This approach offers complementary insights into traditional risk assessment methods, providing clinicians with a novel tool for individualized risk evaluation and decision-making.

Keywords: pulmonary arterial hypertension (PAH); machine learning; XGBoost; SHAP; mortality prediction; personalized medicine; risk stratification

1. Introduction

Pulmonary arterial hypertension (PAH) is a rare, multifactorial cardiovascular disease characterized by increased resistance of the pulmonary arteries due to the proliferation of vascular cells. Over time, the elevated pressure in the pulmonary circulation leads to cardiac overload, remodeling, and ultimately right heart failure, with significantly increased risk of mortality [1].

Early diagnosis and intervention are critical for enhancing patients' prognosis and survival. Predicting unplanned deaths in particular patients suffering from PAH is a significant challenge.

The application of machine learning (ML) models allows for the early identification of patients at increased risk of death, enabling better management of their therapy and more effective medical interventions. Importantly, for the healthcare system, this also aids in optimizing the costs of patient care. ML models can analyze complex relationships between various risk factors, such as functional test results, medical history, pharmacotherapy, and patient lifestyle, providing satisfactory predictions.

Recent years have witnessed a rapid expansion in the use of artificial intelligence (AI) and ML across various domains of cardiology, including risk stratification, arrhythmia detection, and outcome prediction. Large-scale studies have demonstrated that ML-based approaches can outperform traditional statistical methods in predicting adverse events and mortality, especially in complex, multifactorial diseases such as PAH. The integration of AI-driven models into clinical workflows enables the synthesis of high-dimensional data from electronic health records, imaging, and laboratory results, facilitating personalized medicine and more precise prognostication. The utilization of artificial intelligence in cardiology spans multiple aspects. This includes image processing, natural language processing, and wearable devices used by patients daily [2–6]. This trend is driven by both the desire of patients and physicians to leverage the latest technologies and the academic community's pursuit of current trends.

In this context, our study seeks to bridge this gap by leveraging state-of-the-art ML techniques to develop a predictive model for mortality in adult PAH patients. By combining clinical expertise with advanced data analytics, we aim to contribute to the growing body of evidence supporting the use of AI in cardiovascular medicine and to provide clinicians with practical tools for improving patient outcomes.

The aim of this research is to identify risk factors associated with mortality among PAH patients and to develop a machine learning model capable of accurately predicting death. As a tool, it will support clinicians in making informed decisions, optimizing treatment strategies, and ultimately improving patient care and outcomes. According to our current knowledge, the application of ML specifically to mortality prediction in adult PAH patients remains limited in the literature.

2. Materials and Methods

The study utilized the Database of Pulmonary Hypertension in the Polish population (BNP-PL) registry, which contains examination results of PAH patients collected from medical centers across Poland. BNP-PL is the first multicenter prospective registry of adult and pediatric [7] patients with PAH and patients with Chronic Thromboembolic Pulmonary Hypertension (CTEPH) [8] created in Central and Eastern European countries [9]. The methodology of BNP-PL has been previously described [10].

The data obtained from the BNP-PL registry were initially cleaned, and subsequently, 165 features were extracted for 1755 adult patients with PAH. From the baseline visit, information on height, weight, blood pressure, and demographic details was selected. Additionally, detailed data from laboratory tests, right heart catheterization, functional tests, and echocardiography were extracted. An important element was the collected medical history, which included information on current comorbidities, ongoing treatments, and past medical procedures. From the follow-up part of the database, information on death due to sudden cardiac death, respiratory failure, acute pulmonary embolism, acute coronary syndrome, and progression of heart failure from baseline assessment to the first follow-up was obtained.

The initially selected data was then preprocessed by removing extreme outliers and duplicates.

As part of feature engineering, the stroke volume index and the tricuspid annular plane systolic excursion (TAPSE) to systolic pulmonary arterial pressure (sPAP) ratio were calculated. Additionally, the number of diuretics taken was encoded.

All categorical variables were encoded using one-hot encoding, except for the variable "Use of diuretics," whose value reflects the number of different diuretic medications used concurrently. The diuretics used in the BNP-PL population included furosemide, torsemide, thiazide diuretics, and aldosterone antagonists. The age of patients was binned

into 5-year intervals and ordinally encoded. The target variable, which was mortality, was binarily encoded. Next, feature selection was performed using the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm, followed by sequential feature elimination (SFE). While each method independently identifies relevant predictors, their consecutive use may appear methodologically redundant. The rationale for this approach was to leverage the strengths of LASSO in penalizing less informative features and SFE in iteratively refining the feature set based on model performance. Specifically, LASSO was first used to reduce the initial feature pool by discarding 140 features with coefficients shrunk to zero. Subsequently, SFE further eliminated 8 features by evaluating their impact on classification accuracy. This two-step process aimed to enhance model generalizability and interpretability by ensuring that only the most robust predictors were retained. The selected dataset was then split into train and test sets using StratifiedKFold in an 80:20 ratio to ensure that both sets maintained the same class distribution, which is particularly important for imbalanced datasets. An 80:20 split ratio was chosen to maximize the amount of data available for training the model [11].

The resulting datasets had target variable ratios of 1303:101 and 326:25 for the train and test sets, respectively. The test set was reserved for the final evaluation of the model's quality and was not used during the training, validation, and optimization processes of the trained model.

To prepare the machine learning model, tree-based classifiers were planned for use due to their ease of interpretation, which is crucial for the potential use of the results as a clinical support tool for specialist doctors. The eXtreme Gradient Boosting (XGBoost 2.1.3) algorithm is a powerful and efficient implementation of the gradient boosting framework, widely used for supervised learning tasks such as classification and regression [12]. It has gained popularity due to its high performance, scalability, and flexibility. It has been successfully applied in numerous data science competitions and real-world applications, demonstrating its effectiveness and versatility. XGBoost was chosen for its ability to handle missing data, computational speed, ease of optimization, and the possibility of explaining predictions using the Shapley Additive Explanations (SHAP) values [13,14]. The entire training process was conducted using RepeatedStratifiedKFold cross-validation ($n_splits = 10$, $n_repeats = 5$) to ensure robustness and accuracy of the results. Hyperparameter tuning was conducted using the HyperOpt framework with cross-validation, the same as in the training process [15]. During hyperparameter optimization using the hyperopt library, the following hyperparameters were tuned: `colsample_bytree`, `gamma`, `learning_rate`, `max_depth`, `n_estimators`, `min_child_weight`, `reg_alpha`, `reg_lambda`, `scale_pos_weight`, and `subsample`. Due to the high class imbalance in the training dataset, various over-sampling methods such as SMOTE and ADASYN, as well as the combined over- and under-sampling method SMOTETomek, were tested [16]. However, these approaches did not yield satisfactory results, so only hyperparameter optimization was ultimately used.

The fine-tuned and trained model was used to make predictions on the test set, the results of which are presented later in this work. SHAP values were used to achieve explainable artificial intelligence (XAI). SHAP values are based on the concept of Shapley values from cooperative game theory, measuring the contribution of each feature to the model's prediction by considering all possible combinations of features. Using this interpretative technique allows for the identification of the most important features, thereby indicating risk factors both globally and for each individual patient case.

The entire workflow is presented in Figure 1.

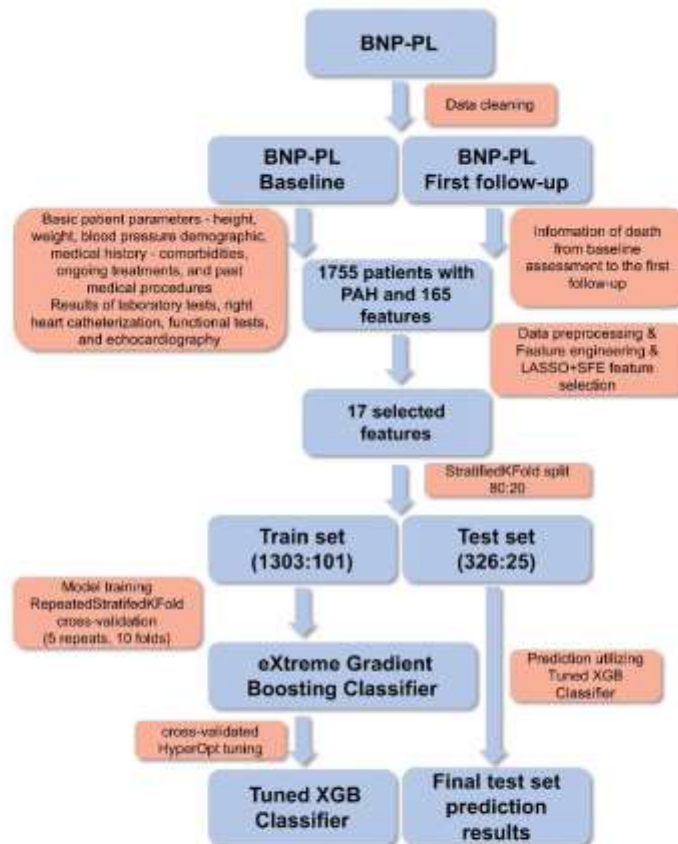


Figure 1. Study workflow. BNP-PL—Database of Pulmonary Hypertension in the Polish population; LASSO—Least Absolute Shrinkage and Selection Operator regression; PAH—Pulmonary Arterial Hypertension; SFE—Sequential Feature Elimination; XGB—eXtreme Gradient Boosting.

To evaluate the quality of the model, typical metrics commonly used for binary classification assessment and the confusion matrix were employed.

The confusion matrix is a summary table of the classification results of the model, containing:

- TP (True Positive): Correctly predicted positive cases.
- TN (True Negative): Correctly predicted negative cases.
- FP (False Positive): Incorrectly predicted positive cases.
- FN (False Negative): Incorrectly predicted negative cases.

These metrics provide a comprehensive view of the model's performance.

Accuracy represents the percentage of correctly classified samples out of all samples. It indicates the overall quality of the model and is very intuitive to interpret. However, for imbalanced datasets, accuracy can be misleading, as it may not reflect the model's performance on the minority class.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Specificity measures the ability to correctly classify the negative class. It is important in scenarios where the cost of Type I errors (false positives) is high and needs to be minimized. High specificity ensures that most actual negative cases are correctly identified by the model.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

Sensitivity, also known as recall, measures the ability to correctly classify the positive class. It is crucial in scenarios where the cost of Type II errors (false negatives) is high and needs to be minimized. High sensitivity ensures that most actual positive cases are identified by the model.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

The Matthews Correlation Coefficient (MCC) is a balanced measure of binary classification quality that evaluates all four values of the confusion matrix. It is particularly reliable for imbalanced datasets. The MCC value ranges from -1 to $+1$, where $+1$ indicates a perfect prediction, 0 indicates no better than random prediction, and -1 indicates total disagreement between prediction and observation.

$$\text{MCC} = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad (4)$$

The Receiver Operating Characteristic—Area Under the Curve (ROC-AUC) is the area under the curve that represents the relationship between sensitivity and 1-specificity at various decision thresholds. It evaluates the model's ability to distinguish between classes; the higher the value, the better the model's ability to predict the positive class.

For the purpose of the present analysis, we defined extended hospitalization for patients with PAH, which refers to hospital stays longer than the average duration for patients with this condition. This duration can vary between hospitals contributing data to the database due to differences in patient populations with the same diagnosis.

All calculations were performed in a Python 3.10.12 environment using publicly available libraries. The computations were conducted on a computer with a CPU featuring 24 threads at 4.5 GHz and 64 GB of RAM. Packages versions: pandas (2.1.4), numpy (1.26.4), hyperopt (0.2.7), shap (0.46.0), scikit-learn (1.4.2), xgboost (2.1.3).

Categorical variables were expressed as counts and compared using the chi-square test. Continuous variables were presented as mean $\pm$ standard deviation (SD) or median and first and third quartiles (Q1, Q3). Continuous variables were analyzed using the U-Mann-Whitney test. A two-tailed test result with a p -value < 0.05 was considered statistically significant.

3. Results

The general characterization of patients with PAH in the BNP-PL registry was presented previously by Kopeć et al. [9]. Between 1 March 2018 and 31 Aug 2023, a total of 1755 PAH patients were enrolled including 902 newly diagnosed and 853 previously

(before 1 March 2018). At the same time, 126 patients died. The median/mean time from baseline visit to death was 10 and 12 months respectively. The study cohort included a diverse range of PAH etiologies. The most common type was idiopathic non-reactive PAH (750 patients, 8% non-survivors), followed by PAH associated with congenital heart disease (483 patients, 5% non-survivors) and PAH associated with connective tissue diseases (305 patients, 10% non-survivors). Less frequent types included idiopathic reactive PAH (121 patients, 2% non-survivors), PAH associated with portal hypertension (45 patients, 7% non-survivors), heritable PAH (19 patients, 5% non-survivors), PAH associated with HIV infection (10 patients, 0% non-survivors), and PAH induced by drugs or toxins (14 patients, 14% non-survivors). Pulmonary veno-occlusive disease was rare (6 patients) but had the highest mortality rate (33%). Only two cases were classified as unknown etiology. The baseline characteristics of patients with PAH were compared in relation to mortality due to sudden cardiac death, respiratory failure, acute pulmonary embolism, acute coronary syndrome, and progression of heart failure in the PAH patient population.

Table 1 below summarizes the statistics of continuous and categorical variables selected by the LASSO regression and SFE used to build the XGB classification model.

Table 1. Baseline characteristics of PAH patients in the BNP-PI registry in relation to survival status. The presented variables were chosen in the feature selection step.

Variable	Survivors (n = 1629)	Non-Survivors (n = 126)	p-Value
Age [years], median (Q1,Q3)	59 (40, 70)	68 (57.75, 74)	0.0000
Gender, male, n (%)	509 (31.25%)	51 (40.48%)	0.0411
Type of PAH: associated with congenital heart disease, n (%)	485 (28.12%)	25 (19.84%)	0.0574
Chronic kidney disease, n (%)	274 (16.82%)	45 (35.71%)	0.0000
Diabetes, n (%)	322 (19.77%)	45 (35.71%)	0.0000
Previous myocardial infarction, n (%)	78 (4.79%)	21 (16.67%)	0.0000
Hospitalized for PAH in the last year, n (%)	240 (15.81%)	37 (31.36%)	0.0000
Extended hospitalization history, n (%)	140 (8.59%)	21 (16.67%)	0.0042
Syncope in the last year, n (%)	97 (6.39%)	18 (15.25%)	0.0006
Symptoms of right heart failure, n (%)	493 (32.48%)	75 (63.56%)	0.0000
Right axis deviation in ECG, n (%)	626 (38.43%)	69 (54.76%)	0.0004
Presence of pericardial effusion, n (%)	290 (18.07%)	39 (33.33%)	0.0001
TAPSE/sPAP ratio [mm/mmHg], median (Q1,Q3)	0.25 (0.18, 0.38)	0.21 (0.15, 0.28)	0.0006
Distance in 6-min walk test [m], median (Q1,Q3)	360 (240, 460)	231 (120, 322.5)	0.0000

Table 1. Cont.

Variable	Survivors (n = 1629)	Non-Survivors (n = 126)	p-Value
Plasma NT-proBNP concentration [pg/mL], median (Q1,Q3)	813.9 (250.55, 2482.5)	2983.5 (1445.75, 5529.75)	0.0000
Use of diuretics, [0/1/2/3/4] n (%)	487 (29.9%)	11 (8.73%)	0.0000
	563 (34.56%)	46 (36.51%)	
	461 (28.3%)	53 (42.06%)	
	109 (6.69%)	12 (9.52%)	
Parenteral use of Trepresinil or Epoprostenol, n (%)	9 (0.55%)	4 (3.17%)	0.0000
	281 (17.25%)	49 (38.89%)	

The classification model based on the XGBoost algorithm demonstrates moderate performance on the test data. While the model maintains good sensitivity and a low rate of false negatives, it exhibits a relatively high proportion of false positive predictions, which translates into low precision. Table 2 below summarizes the key performance metrics for the model. These metrics indicate the model's effectiveness in predicting death by the next follow-up in the population of Polish patients with PAH.

Table 2. Mean performance metrics for the test set.

Metric	Result	95% Confidence Interval
Accuracy	0.738	0.695–0.783
Sensitivity	0.800	0.636–0.947
Specificity	0.733	0.688–0.779
AUC	0.286	0.199–0.399
ROC-AUC	0.767	0.682–0.843

Particular attention should be given to the sensitivity of the predictive model, which, considering the significant class imbalance, was made possible through the optimization parameters available for the XGBoost algorithm. Table 3 contains the confusion matrix for the test set.

Table 3. Confusion matrix for the test set.

True		Predicted	
		Negative TN = 239 FN = 5	Positive FP = 87 TP = 20
Negative			
Positive			

The satisfactory results presented by performance metrics indicate that the use of the XGBoost algorithm to predict death by the next follow-up in the population of Polish patients with PAH is fully justified. The features indicated by SHAP values highlight significant clinical parameters that can be used for risk assessment and modification of patient treatment. The value of feature engineering in improving the model's quality is noteworthy, as evidenced by the high SHAP values for features prepared based on domain knowledge. The SHAP “beeswarm” plot represents the distribution of SHAP values for each feature, showing the variability in feature impact across different instances. This visualization helps to understand how different feature values influence the model's predictions (Figure 2). The features identified by the model are reflected in the ESC guidelines for stratifying patients into risk groups, including blood NT-proBNP concentration, 6-min walk test result, TA/PA ratio to sPAP ratio [17,18], and the presence of pericardial effusion. All these parameters are considered to have the greatest impact on predictive ability, as expressed by SHAP values.

Locally indicating the most important features allows for easy interpretation and potential medical intervention (Figure 3). Among the features that proved to be most significant for the predictive model were symptoms of right ventricular failure, TAPSE to sPAP ratio, 6MWT, presence of pericardial fluid, NT-proBNP serum concentration, right axis deviation in ECG, age, gender, comorbidities—diabetes and chronic kidney disease—and medications used in PAH therapy or to treat comorbidities. These medications include prostacyclin analog treprostinil and epoprostenol, diuretics such as furosemide and aldosterone antagonists or thiazide diuretics.

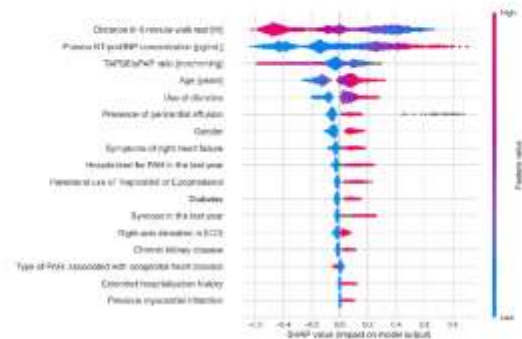


Figure 2. Distribution of SHAP values for all 17 features. The color gradient indicates the actual feature values, with blue representing lower values and red representing higher values, and gray representing missing values. The horizontal spread of points for each feature indicates the variability in the SHAP values, reflecting the feature's influence on the model's predictions. A wider vertical spread of points for a feature suggests a larger number of observations or greater diversity in the feature's values.

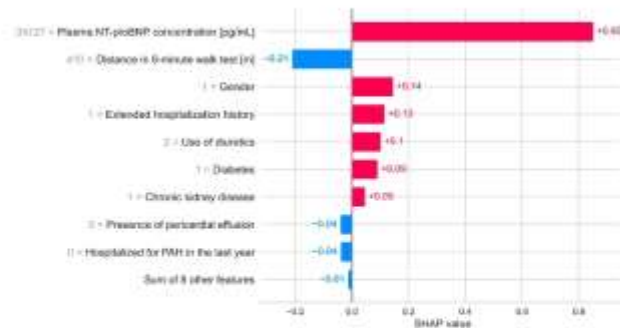


Figure 3. The bar plot of the top nine features illustrates the SHAP values for a specific patient, highlighting the contribution of each feature to the model's prediction. Each bar represents a feature,

with the length of the bar indicating the magnitude of the feature's impact on the prediction. Positive SHAP values (bars extending to the right) indicate features that increase the risk of death, while negative SHAP values (bars extending to the left) indicate features that decrease the risk of death. The label on the Y-axis corresponds to the actual values of the features for this patient. This detailed breakdown allows for a clear understanding of the key factors driving the model's decision for this particular patient. In this case, the model predicted the patient as a non-survivor, and indeed, the patient belonged to the non-survivor group.

4. Discussion

The obtained results provide hope for the development of a tool for cardiologists that allows for the optimization of treatment for patients with PAH. Such an approach may not only enable personalized treatment but also potentially improve patients' prognosis [19,20].

The most closely related study in terms of its assumptions, predicting 30-day readmission in pediatric pulmonary hypertension, utilized the CatBoost algorithm, which belongs to the group of gradient boosting algorithms [21]. The authors of the study achieved very good results with an AUC of 0.81, high accuracy of 0.74, sensitivity of 0.78, and specificity of 0.74. However, due to the different patient population and the limitation of prediction to 30-day readmission, direct comparison of the obtained results is not possible. Another example of a related study is the research that utilized machine learning to predict one-year mortality in newly diagnosed patients with PAH. The authors, using data from cardiac magnetic resonance imaging, achieved very satisfactory results, improving the REVEAL score predictions from a c-index of 0.71 to 0.76. However, fundamental differences in methodological approach and the type of data used do not allow for a direct comparison of the results obtained in this study with those of the present work [22]. The lack of studies with assumptions more closely aligned to those presented in this research does not allow for comparison of the results obtained by other researchers [23–26]. Moreover, the recent literature highlights the successful application of artificial intelligence in various fields of cardiology, including arrhythmia management and cardiac electrophysiology. For example, Cersosimo et al. demonstrated that AI-based models, such as ChatGPT-4o, can support ECG interpretation, arrhythmia detection, procedural guidance during ablation, and risk stratification for sudden cardiac death. These advances illustrate the transformative potential of AI in improving diagnostic accuracy, optimizing clinical workflows, and personalizing therapy. However, the authors also emphasize the importance of balancing AI adoption with clinical expertise and ethical considerations, underscoring the need for collaboration between clinicians and AI developers to ensure reliable and transparent solutions in cardiovascular care [27].

A comparative analysis was performed between gradient boosting frameworks such as LightGBM, CatBoost, and AdaBoost, alongside XGBoost. Although LightGBM and CatBoost are known for their computational efficiency and advanced handling of categorical variables, and AdaBoost for its simplicity, our experiments showed that these algorithms consistently yielded inferior predictive performance compared to XGBoost on our dataset. Additionally, given the relatively small size of our dataset, differences in computation time between these frameworks were negligible. To maintain clarity and focus, only XGBoost results were included in the final manuscript. The choice of XGBoost was further justified by its robust regularization capabilities, scalability, and proven effectiveness in handling clinical tabular data, which are critical for PAH mortality prediction tasks. This approach aligns with current literature, where XGBoost remains a leading choice for biomedical machine learning applications.

It is important to emphasize that the SHAP values indicating the significance of medications used by patients do not point to the medication itself, which is used according

to the treatment guidelines for PAH patients. Instead, they reflect the clinical condition of the patient that necessitates appropriate pharmacological treatment. The presented model serves as a classifier indicating the likelihood of a patient's mortality before the next follow-up, rather than specifying the exact therapeutic steps to avoid such outcomes. The identification of medications as significant predictors does not imply a need to alter their usage but highlights the necessity for closer monitoring of patients' health status, as clinical deterioration may occur independently of the medication itself (e.g., diuretics). Moreover, parenteral prostacyclins such as treprostinil or epoprostenol are typically reserved for patients with more advanced disease, serving as an indicator of baseline severity rather than a direct contributor to increased mortality risk [26]. Conversely, medications that modulate heart rate, such as beta-blockers or ivabradine, which are used to counteract hyperadrenergic drive in heart failure, may adversely affect survival in PAH and are generally contraindicated unless justified by specific comorbidities [29]. The presence of diabetes among the highly significant features is quite understandable. Patients with PAH and diabetes present with more advanced pulmonary vascular disease and worse survival rates compared to their counterparts without diabetes [30].

The demonstrated significant features may require confirmation in clinical studies and serve as a kind of guideline for what should be investigated in multicenter clinical trials, especially in the context of increased risk of readmission or death. Many of the predictive factors indicated by the model are complex and multifactorial, making them difficult to interpret unequivocally. For example, prolonged hospitalizations, where the reason for the extension is unknown, cannot be conclusively attributed solely to the underlying disease, PAH. It might have been influenced by factors independent of the patient, such as the occurrence of an infection or an injury that required the patient to remain under observation in the hospital. The prepared model has its limitations due to the available data. It would be necessary to train on a larger dataset, preferably containing information about patients from multiple populations. A dataset containing data only from one country may be biased by the healthcare funding system, which affects the medications used, significantly impacting the model's predictive capabilities. The limitations in the number of patients in the database and bias towards a specific population also affect the model's stability and the risk of poor prediction on entirely external datasets. Another problem is the unequal time from baseline assessment to follow-up among patients. Introducing a fixed time interval for follow-up registration would significantly improve the applicability of machine learning.

Most current research focuses on diagnostic algorithms in PAH. Nonetheless, technologies such as gradient boosting and feature importance explanation using SHAP are also prevalent in these studies [31–34]. The utilization of explainable machine learning models appears to be essential in the field of medicine [35].

Additionally, there is a noticeable need for the creation of multidisciplinary research teams where clinicians can provide their domain expertise, and data scientists can use this knowledge to build better models that yield satisfactory results and are more robust. The development of a personalized approach to patient treatment in the future will require us to utilize modern technologies based on artificial intelligence, whose advancement has significantly accelerated in recent years [36].

Despite satisfactory results, the presented machine learning model still leaves a considerable margin of error, especially when viewed through the lens of human health and life. One of the next steps in the development of similar studies should be to investigate the impact of PAH type on the predictive capabilities of the model. The presented model has considerable issues with precision, resulting in a high rate of false positive predictions, as shown in Table 3. However, we believe that drawing attention to a patient whose actual risk of death is lower is a lesser problem than missing a patient with a genuinely high risk

of death. From an ethical and moral standpoint, this approach seems more appropriate. Certainly, the next stage of our future work will be to search for a model that maintains both a low rate of false positive and false negative predictions. However, to achieve this, it is necessary to first address the remaining limitations. AI solutions facilitate and accelerate the work of physicians, but decision-making should always be critically evaluated through the lens of clinical experience [37].

5. Conclusions

The application of machine learning to predict mortality in patients with PAH yields satisfactory results. However, further research based on large international registries of this rare disease is required.

In the future, a tool similar to this could facilitate work in every hospital and provide invaluable support in the daily challenges of caring for patients with PAH. The integration of AI-driven predictive analytics in PAH management holds immense promise for revolutionizing clinical decision-making and improving patient outcomes. By leveraging advanced machine learning algorithms and comprehensive patient data, healthcare providers can identify high-risk individuals, customize treatment approaches, and implement proactive interventions to prevent mortality.

Author Contributions: Conceptualization, L.L. and G.G.; writing—original draft preparation, L.L. and G.G.; writing—review and editing, G.G., M.Z., M.W., M.K., T.M.-K., A.S.-G., I.S., I.C., P.B., M.J., B.K.-D., K.P., K.M.-S., E.M. (Ewa Malinowska), M.P.-P., F.L., M.T., W.J., M.F., E.M. (Ewa Mruczek), Z.G., A.P., K.B.-L., P.P., O.D.-D., K.W., J.W.-J. and G.K.; visualization, L.L.; supervision, G.G., G.K. and M.W.; project administration, G.G.; funding acquisition, G.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study protocol adhered to the ethical guidelines of the Declaration of Helsinki and was reviewed and accepted by the local bioethical committee of the Bioethics Committee of the Regional Medical Chamber in Kraków, which confirmed that, due to the purely statistical and non-interventional nature of the study, no formal ethical approval was required (Decision No. OIL/KBL/27/2018, dated 17 July 2018).

Informed Consent Statement: Patients provided informed consent upon enrollment in the BNP-PL registry.

Data Availability Statement: The data underlying this article will be shared on reasonable request to the BNP-PL steering committee.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
BNP-PL	Database of Pulmonary Hypertension in the Polish population
CTEPH	Chronic Thromboembolic Pulmonary Hypertension
LASSO	Least Absolute Shrinkage and Selection Operator algorithm
MCC	Matthews Correlation Coefficient
ML	Machine learning
PAH	Pulmonary Arterial Hypertension
ROC-AUC	The Receiver Operating Characteristic—Area Under the Curve
SFE	Sequential Feature Elimination
SHAP	Shapley Additive Explanations
sPAP	Systolic Pulmonary Arterial Pressure

TAPSE	Tricuspid Annular Plane Systolic Excursion
XAI	eXplainable Artificial Intelligence
XGBoost	eXtreme Gradient Boosting

References

- Humbert, M.; Kovacs, G.; Hoeper, M.M.; Badagliacca, R.; Berger, R.M.B.; Brida, M.; Carlson, J.; Coats, A.J.S.; Escribano-Salas, P.; Ferrari, P.; et al. 2022 ESC/ERS Guidelines for the Diagnosis and Treatment of Pulmonary Hypertension. *Eur. Heart J.* **2022**, *43*, 3618–3731. [\[CrossRef\]](#)
- Johnson, K.W.; Torres-Soto, J.; Glicksberg, B.S.; Shameer, K.; Miotto, R.; Ali, M.; Ashley, E.; Dudley, J.T. Artificial Intelligence in Cardiology. *J. Am. Coll. Cardiol.* **2018**, *71*, 2668–2679. [\[CrossRef\]](#)
- Watson, X.; D'Souza, J.; Cooper, D.; Markham, R. Artificial Intelligence in Cardiology: Fundamentals and Applications. *Intern. Med. J.* **2022**, *32*, 912–920. [\[CrossRef\]](#) [\[PubMed\]](#)
- Weston, S.J.; Ichter, J.E. Artificial Intelligence and Machine Learning in Cardiology. *JACC Cardiovasc. Interv.* **2019**, *12*, 1312–1314. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cohen-Shelly, M.; Atia, Z.I.; Friedman, P.A.; Ito, S.; Tassayagh, B.A.; Ko, W.Y.; Murphree, D.H.; Michelena, H.I.; Enriquez-Sarano, M.; Carter, R.E.; et al. Electrocardiogram Screening for Aortic Valve Stenosis Using Artificial Intelligence. *Eur. Heart J.* **2021**, *42*, 2885–2896. [\[CrossRef\]](#)
- Piescho, K.; Hiczkiewicz, J.; Budzianowski, J.; Musielak, B.; Hiczkiewicz, D.; Faron, W.; Rzednicki, J.; Burekhardt, P. Clinical Applications of Artificial Intelligence in Cardiology on the Verge of the Decade. *Cardiol. J.* **2021**, *28*, 460–472. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kwiatkowska, J.; Zuk, M.; Migdal, A.; Kusa, J.; Soba, E.; Zygiel, K.; Przetočka, K.; Werynski, P.; Baraszkas, P.; Rzeźnicki-Bieniaszewska, A.; et al. Children and Adolescents with Pulmonary Arterial Hypertension: Baseline and Follow-Up Data from the Polish Registry of Pulmonary Hypertension (BNP-PL). *J. Clin. Med.* **2020**, *9*, 1717. [\[CrossRef\]](#)
- Bodasiewicz, M.; Kurzyńska, M.; Knap, G.; Pruszczyk, P.; Mroczek, E.; Mulałek-Kubzdela, T.; Skoczylas, I.; Błaszcak, P.; Chrzanowski, L.; Jagasiewicz, M.; et al. Hemodynamic, Echocardiographic, and Demographic Profiles of Patients with Chronic Thromboembolic Pulmonary Hypertension and Atrial Fibrillation: A Multicenter Cohort Study. *J. Heart Lung Transplant.* **2025**, *44*, 1239–1248. [\[CrossRef\]](#)
- Kopeć, G.; Kurzyńska, M.; Mroczek, E.; Chrzanowski, L.; Mulałek-Kubzdela, T.; Skoczylas, I.; Kusiński, B.; Pruszczyk, P.; Błaszcak, P.; Lewicka, E.; et al. Characterization of Patients with Pulmonary Arterial Hypertension: Data from the Polish Registry of Pulmonary Hypertension (BNP-PL). *J. Clin. Med.* **2020**, *9*, 173. [\[CrossRef\]](#)
- Kopeć, G.; Kurzyńska, M.; Mroczek, E.; Chrzanowski, L.; Mulałek-Kubzdela, T.; Skoczylas, I.; Kusiński, B.; Pruszczyk, P.; Błaszcak, P.; Lewicka, E.; et al. Database of Pulmonary Hypertension in the Polish Population (BNP-PL): Design of the Registry. *Kardiol. Pol.* **2019**, *77*, 972–974. [\[CrossRef\]](#)
- Raschka, S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *arXiv* **2020**, arXiv:1811.12808. [\[CrossRef\]](#)
- Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.
- Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3146–3154.
- Lundberg, S.M.; Fries, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [\[CrossRef\]](#)
- Bergstra, J.; Yamins, D.; Cox, D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; Dasgupta, S., McAllester, D., Eds.; PMLR: Atlanta, GA, USA, 2013; Volume 28, pp. 115–123.
- Aggarwal, R.; Menzies, J.A.; Nyarko, F.; Arku, D.; Mbebe-Baiden, B.; Opoku, E.; Noye-Nortey, E.N. Addressing Class Imbalance Problem in Health Data Classification: Practical Application From an Oversampling Viewpoint. *Appl. Comput. Intell. Soft. Comput.* **2025**, *2025*, 1015769. [\[CrossRef\]](#)
- Coladillo, A.; Hoffmann-Vold, A.-M.; Pellicano, C.; Romanelli, A.; Gabrielli, A.; Hachulla, E.; Smith, V.; Simeón-Aznar, C.-J.; Castellvi, I.; Airò, P.; et al. The Role of TAPSE/SPAP Ratio in Predicting Pulmonary Hypertension and Mortality in the Systemic Sclerosis EUSTAR Cohort. *Autoimmun. Rev.* **2023**, *22*, 103290. [\[CrossRef\]](#)
- Palacios-Moguel, P.; Cueto-Robledo, G.; González-Pacheco, H.; Ortega-Hernández, J.; Torres-Rojas, M.B.; Nava-Vergara, D.I.; García-Cosío, M.; González-Nájera, C.A.; Narváez-Orián, C.A.; Sandoval, J. The Role of the TAPSE/SPAP Ratio as a Predictor of Mortality in Pulmonary Arterial Hypertension: Its Value for Patient Risk Stratification. *JHLT Open* **2025**, *7*, 100168. [\[CrossRef\]](#)

19. Jain, S.; Kaur, A.; Qadoor, A.; Ghosh, V.; Thota, S.; Banala, M.; Lee, J.; Yerrapragada, G.; Elangovan, P.; Shariff, M.N.; et al. Leveraging Artificial Intelligence for the Diagnosis of Systemic Sclerosis Associated Pulmonary Arterial Hypertension: Opportunities, Challenges, and Future Perspectives. *Adv. Respir. Med.* **2025**, *93*, 47. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Benesch Vidal, M.L.; Arvanitaki, A.; Diller, G.P. The Role of Artificial Intelligence and Mobile Health in Diagnosis and Management of Pulmonary Arterial Hypertension. *Int. J. Cardiol. Congest. Heart Dis.* **2025**, *22*, 100622. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Duan, M.; Shu, T.; Zhao, B.; Xiang, T.; Wang, J.; Huang, H.; Zhang, Y.; Xiao, P.; Zhou, B.; Xie, Z.; et al. Explainable Machine Learning Models for Predicting 30-Day Readmission in Pediatric Pulmonary Hypertension: A Multicenter, Retrospective Study. *Front. Cardiovasc. Med.* **2022**, *9*, 919224. [\[CrossRef\]](#)
22. Alabed, S.; Uthoff, J.; Zhou, S.; Gang, P.; Dwivedi, K.; Alandijani, F.; Gerdling, R.; Scheib, L.; Brook, M.; Shahin, Y.; et al. Machine Learning Cardiac-MRI Features Predict Mortality in Newly Diagnosed Pulmonary Arterial Hypertension. *Eur. Heart J.-Digit. Health* **2022**, *3*, 266–275. [\[CrossRef\]](#)
23. Attarpour Esfahani, S.; Babar Ali, N.; Farina, J.M.; Scalia, I.G.; Pereyra, M.; Abbas, M.T.; Javadi, N.; Blamee, N.N.; Abdelfattah, F.E.; Awad, K.; et al. A Comprehensive Review of Artificial Intelligence (AI) Applications in Pulmonary Hypertension (PH). *Medicina* **2025**, *61*, 85. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Salehi, M.; Alabed, S.; Sharkey, M.; Meier, A.; Dwivedi, K.; Yardibi, T.; Seiej, M.; Hameed, A.; Charalampopoulos, A.; Kiely, D.G.; et al. Artificial Intelligence-Based Echocardiography Assessment to Detect Pulmonary Hypertension. *ERJ Open Res.* **2025**, *11*, 00392–2024. [\[CrossRef\]](#)
25. Kwock, J.; Kim, K.-H.; Medina-Inojosa, J.; Jeon, K.-H.; Park, J.; Oh, B.-H. Artificial Intelligence for Early Prediction of Pulmonary Hypertension Using Electrocardiography. *J. Heart Lung Transplant.* **2020**, *39*, 805–814. [\[CrossRef\]](#)
26. Celestin, B.; Bagherzadeh, S.P.; Santana, E.; Frost, M.; Iversen, M.; Hermanson, E.N.; Sweet, A.; Zacharian, R.T.; Hummel, Y.M.; Rendoo, C.G.; et al. Artificial Intelligence-Based Echocardiography in Pulmonary Arterial Hypertension. *Chest* **2025**, in press. [\[CrossRef\]](#)
27. Cersosimo, A.; Zito, E.; Pierucci, N.; Materucci, A.; La Fazio, V.M. A Talk with ChatGPT: The Role of Artificial Intelligence in Shaping the Future of Cardiology and Electrophysiology. *J. Pers. Med.* **2025**, *15*, 205. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Saleh, K.M.; Mallat, J.; Mohammed, S.; Bodi, G.; Alkazazi, H.; Salim, S.; Elhennawi, M.; Isqbal, T.; Sabbour, H. Comparative Efficacy and Safety of Prostacyclin Therapies for Pulmonary Arterial Hypertension: A Systematic Review and Network Meta-Analysis. *Front. Med.* **2025**, *12*, 1643220. [\[CrossRef\]](#)
29. Kaemmerer, H.; Diller, G.P.; Dübner, I.; Achenbach, S.; Eichstaedt, C.A.; Eicken, A.; Freilinger, A.; Freilinger, S.; Geiger, R.; Gorenflo, M.; et al. Pulmonary Hypertension in Adults with Congenital Heart Defects (ACHDs) in Light of the 2022 ESC PAH Guidelines—Part II: Supportive Therapy, Special Situations (Pregnancy, Contraception, Non-Cardiac Surgery), Targeted Pharmacotherapy, Organ Transplantation, Special Management (Shunt Lesion, Left Ventricular Disease, Univentricular Hearts), Interventions, Intensive Care, ACHD Follow-up, Future Perspective. *Cardiovasc. Diagn. Ther.* **2024**, *14*, 921–934. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Joras, K.; Kurzyba, M.; Mroczek, F.; Chrzaniowski, I.; Mulałek-Kubacka, T.; Skoczylas, J.; Blaszczyk, P.; Grzesik, G.; Misia-Słoc, K.; Kuśmierczyk, B.; et al. Impact of Diabetes Mellitus on Disease Severity and Patient Survival in Idiopathic Pulmonary Arterial Hypertension: Data from the Polish Multicentre Registry (BNP-PL). *Cardiovasc. Diabetol.* **2023**, *22*, 177. [\[CrossRef\]](#)
31. Zheng, H.; Waqar, M.M.; Arif, S.; Sherazi, S.W.A.; Son, S.H.; Lee, J.Y. An Explainable Machine Learning-Based Prediction Model for In-Hospital Mortality in Acute Myocardial Infarction Patients with Typical Chest Pain. In Proceedings of the 2023 The 6th International Conference on Software Engineering and Information Management, Palmerston North, New Zealand, 31 January 2023; ACM: New York, NY, USA, 2023; pp. 44–49.
32. Liao, Z.; Liu, K.; Ding, S.; Zhao, Q.; Jiang, Y.; Wang, L.; Huang, T.; Yang, L.; Luo, D.; Zhang, E.; et al. Automatic Echocardiographic Evaluation of the Probability of Pulmonary Hypertension Using Machine Learning. *Pulm. Circ.* **2023**, *13*, e12272. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Kogan, F.; Didden, F.-M.; Lee, F.; Nuenwe, A.; Stamatidis, D.; Matarazzo, S.; Quinn, D.; Rosenberg, D.; Chehoud, C.; Bridges, C. A Machine Learning Approach to Identifying Patients with Pulmonary Hypertension Using Real-World Electronic Health Records. *Int. J. Cardiol.* **2023**, *374*, 95–99. [\[CrossRef\]](#)
34. Anand, V.; Weston, A.D.; Scott, C.G.; Kane, G.C.; Pellikka, P.A.; Carter, R.E. Machine Learning for Diagnosis of Pulmonary Hypertension by Echocardiography. *Mayo Clin. Proc.* **2024**, *99*, 260–270. [\[CrossRef\]](#)
35. Durán, J.M.; Jongasma, K.R. Who Is Afraid of Black Box Algorithms? On the Epistemological and Ethical Basis of Trust in Medical AI. *J. Med. Ethics* **2021**, *47*, 329–335. [\[CrossRef\]](#)

36. Patrascaru, O.S.; Tutunaru, D.; Masot, C.L.; Dragostin, O.M.; Fulga, A.; Nechita, L.; Ciubara, A.B.; Piraiasu, A.I.; Stamate, E.; Poalelungi, D.G.; et al. Future Horizons: The Potential Role of Artificial Intelligence in Cardiology. *J. Pers. Med.* **2024**, *14*, 656. [[CrossRef](#)] [[PubMed](#)]
37. Rosemann, A.; Zhang, X. Exploring the Social, Ethical, Legal, and Responsibility Dimensions of Artificial Intelligence for Health—A New Column in Intelligent Medicine. *Intell. Med.* **2022**, *2*, 103–109. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.