

Prognozowanie przeżywalności pacjentów z niewydolnością serca

Fabian Żebrowski 

Politechnika Białostocka, Wydział Mechaniczny

e-mail: 95809@student.pb.edu.pl

Małgorzata Zdrodowska 

Politechnika Białostocka, Wydział Mechaniczny

e-mail: m.zdrodowska@pb.edu.pl

DOI: 10.65731/az-2026-0077

Streszczenie

Niewydolność serca stanowi istotny problem kliniczny ze względu na wysoką śmiertelność oraz złożoność czynników wpływających na rokowanie pacjentów. Celem pracy była ocena możliwości wykorzystania metod uczenia maszynowego do przewidywania przeżywalności pacjentów z niewydolnością serca oraz określenie wpływu selekcji cech na skuteczność klasyfikacji. Analizę przeprowadzono na zbiorze Heart Failure Clinical Records obejmującym dane kliniczne 299 pacjentów z dysfunkcją skurczową lewej komory. Do klasyfikacji zastosowano algorytmy Naive Bayes, J48, Random Forest oraz regresję logistyczną, wykorzystując 10-krotną walidację krzyżową. Porównano modele wykorzystujące pełny zestaw atrybutów oraz zestaw cech wybranych za pomocą metody InfoGainAttributeEval z algorytmem Ranker. Najlepsze wyniki dla pełnego zestawu danych uzyskano dla klasyfikatora Random Forest, dla którego TP Rate wyniósł 0,849, Precision 0,847, a ROC Area 0,911. Selekcja cech poprawiła wyniki klasyfikacji dla większości analizowanych algorytmów, przy jednoczesnym ograniczeniu liczby zmiennych wykorzystywanych przez modele. Uzyskane wyniki wskazują, że metody uczenia maszynowego mogą skutecznie wspierać ocenę przeżywalności pacjentów z niewydolnością serca, a odpowiedni dobór cech pozwala ograniczyć złożoność modelu bez istotnego pogorszenia jego zdolności predykcyjnych.

Słowa kluczowe

niewydolność serca, uczenie maszynowe, przeżywalność pacjentów, selekcja cech

Wstęp

Niewydolność serca jest stanem klinicznym, w którym serce nie jest w stanie zapewnić odpowiedniego przepływu krwi, umożliwiającego pokrycie zapotrzebowania metabolicznego organizmu [Narodowy Instytut Kardiologii, b.r.]. Według klasyfikacji New York Heart Association (NYHA) wyróżnia się cztery klasy niewydolności serca, określające stopień nasilenia objawów i ograniczenia aktywności fizycznej pacjenta [American Heart Association, 2025]. Do najczęściej występujących objawów należą duszność, szczególnie podczas wysiłku, kołatanie serca, przewlekłe zmęczenie, obrzęki kostek i podudzi oraz zawroty głowy [Narodowy Instytut Kardiologii, b.r.]. W zaawansowanych stadiach choroby objawy mogą istotnie ograniczać codzienne funkcjonowanie pacjenta, dlatego istotne znaczenie mają wczesna diagnostyka oraz odpowiednio dobrane leczenie.

W Polsce na niewydolność serca choruje około 1,3 mln osób [Narodowy Instytut Kardiologii, 2025]. Liczba nowych przypadków utrzymuje się na wysokim poziomie i szacowana jest na około 130–150 tys. rocznie. Duża liczba pacjentów z niewydolnością serca stanowi istotne wyzwanie dla systemu opieki zdrowotnej, w szczególności w zakresie diagnostyki, leczenia oraz długoterminowego monitorowania chorych. Wspomaganie oceny rokowania pacjentów za pomocą metod analizy danych i uczenia maszynowego może ułatwić identyfikację osób obciążonych zwiększonym ryzykiem zgonu. Modele predykcyjne oparte na danych klinicznych mogą tym samym stanowić narzędzie wspierające personel medyczny w procesie podejmowania decyzji.

Celem pracy była ocena możliwości wykorzystania wybranych metod uczenia maszynowego do przewidywania przeżywalności pacjentów z niewydolnością serca na podstawie danych zawartych w zbiorze Heart Failure Clinical Records [UCI Machine Learning Repository, 2020]. Dodatkowym celem było zbadanie wpływu selekcji cech na skuteczność klasyfikacji oraz porównanie wyników uzyskanych dla pełnego i zredukowanego zestawu atrybutów. Analiza nawiązuje do badań Chicco i Jurmana [Chicco i Jurman, 2020], w których wykazano istotne znaczenie stężenia kreatyniny w surowicy oraz frakcji wyrzutowej w prognozowaniu przeżywalności pacjentów z niewydolnością serca.

1. Przegląd literatury

Niewydolność serca stanowi istotny problem zdrowia publicznego ze względu na dużą liczbę chorych, wysoką śmiertelność oraz częste hospitalizacje. W Polsce z niewydolnością serca żyje około 1,3 mln osób [Narodowy Instytut Kardiologii,

2025]. Szczególnie istotnym problemem pozostaje zaawansowana postać choroby, związana z pogorszeniem jakości życia, częstszymi hospitalizacjami oraz zwiększonym ryzykiem zgonu. Ryzyko wystąpienia niewydolności serca wzrasta wraz z wiekiem, a choroba szczególnie często dotyczy osób starszych.

Duża liczba zachorowań i hospitalizacji związanych z chorobami układu krążenia stanowi istotne obciążenie dla systemu ochrony zdrowia. W odpowiedzi na ten problem w Polsce wdrożono Narodowy Program Chorób Układu Krążenia na lata 2022–2032. Do jego głównych celów należą ograniczenie zachorowalności i umieralności z powodu chorób układu krążenia, zmniejszenie regionalnych różnic w dostępie do diagnostyki i leczenia oraz ograniczenie występowania czynników ryzyka, takich jak nadciśnienie tętnicze, zaburzenia lipidowe, otyłość, cukrzyca i palenie tytoniu [Narodowy Instytut Kardiologii, 2025].

Jednym z istotnych zagadnień związanych z leczeniem pacjentów z niewydolnością serca jest właściwa ocena rokowania. Ze względu na złożony charakter choroby przewidywanie jej przebiegu wyłącznie na podstawie pojedynczych parametrów klinicznych może być niewystarczające. W literaturze wskazuje się również na ograniczenia tradycyjnych metod oceny stanu pacjenta. Przykładowo klasyfikacja NYHA, pomimo powszechnego zastosowania w praktyce klinicznej, charakteryzuje się pewnym stopniem subiektywności i ograniczoną powtarzalnością oceny [Raphael i in., 2007]. Problemem pozostaje także trafne przewidywanie rokowania przez lekarzy, co uzasadnia poszukiwanie dodatkowych metod wspomagających ocenę ryzyka [Buchan i in., 2019].

Rozwój metod analizy danych umożliwił wykorzystanie algorytmów uczenia maszynowego do wspomagania diagnozowania i prognozowania przebiegu chorób układu krążenia. Modele takie pozwalają na jednoczesną analizę wielu parametrów klinicznych oraz identyfikację zależności, które mogą być trudne do określenia za pomocą klasycznych metod analizy danych. W przypadku niewydolności serca metody uczenia maszynowego są wykorzystywane między innymi do przewidywania śmiertelności, ryzyka ponownej hospitalizacji oraz identyfikacji pacjentów należących do grup podwyższonego ryzyka.

Szczególnie istotne dla niniejszej pracy są badania Chicco i Jurmana [Chicco i Jurman, 2020], którzy analizowali dane kliniczne 299 pacjentów z niewydolnością serca. Autorzy zastosowali metody uczenia maszynowego w celu określenia czynników mających największe znaczenie dla przewidywania przeżywalności pacjentów. Wykazali, że spośród analizowanych parametrów szczególnie istotne znaczenie prognostyczne mają stężenie kreatyniny w surowicy oraz frakcja wyrzutowa lewej komory. Wyniki ich badań wskazują, że stosunkowo niewielka liczba odpowiednio

dobrych cech może umożliwić uzyskanie dobrej jakości predykcji przeżywalności pacjentów [Chicco i Jurman, 2020].

Stężenie kreatyniny w surowicy jest parametrem wykorzystywanym w ocenie funkcji nerek, których zaburzenia często współwystępują z niewydolnością serca i mogą wpływać na rokowanie pacjenta. Z kolei frakcja wyrzutowa określa procent objętości krwi wypompowywanej z lewej komory podczas jej skurczu i stanowi jeden z podstawowych parametrów wykorzystywanych w ocenie funkcji skurczowej serca. Znaczenie tych parametrów w prognozowaniu przeżywalności uzasadnia ich uwzględnienie podczas budowy modeli klasyfikacyjnych [Chicco i Jurman, 2020].

Wyniki dotychczasowych badań wskazują zatem na możliwość wykorzystania metod uczenia maszynowego jako narzędzi wspomagających ocenę rokowania u pacjentów z niewydolnością serca. Istotnym zagadnieniem pozostaje jednak dobór odpowiednich cech oraz algorytmu klasyfikacyjnego. Ograniczenie liczby zmiennych wejściowych może zmniejszyć złożoność modelu, ułatwić jego interpretację oraz skrócić czas obliczeń, pod warunkiem zachowania odpowiedniej jakości klasyfikacji. Z tego względu w niniejszej pracy porównano działanie kilku algorytmów uczenia maszynowego przed i po przeprowadzeniu selekcji cech.

2. Metodyka badań

Do przeprowadzenia badań wykorzystano zbiór Heart Failure Clinical Records, opracowany przez Chicco i Jurmana i udostępniony w repozytorium UCI Machine Learning Repository [UCI Machine Learning Repository, 2020]. Zbiór zawiera dane kliniczne 299 pacjentów, w tym 105 kobiet i 194 mężczyzn, w wieku od 40 do 95 lat. Dane zostały zgromadzone w Institute of Cardiology and Allied Hospital w Faisalabadzie w Pakistanie w okresie od kwietnia do grudnia 2015 roku. Badaniem objęto pacjentów z dysfunkcją skurczową lewej komory, należących do III lub IV klasy niewydolności serca według klasyfikacji NYHA.

Zbiór danych obejmuje 13 atrybutów opisujących stan kliniczny pacjentów oraz przebieg obserwacji. Wśród nich znajdują się zarówno zmienne ilościowe, takie jak wiek, stężenie kreatyniny i sodu w surowicy, frakcja wyrzutowa oraz liczba płytek krwi, jak również zmienne binarne określające między innymi występowanie anemii, cukrzycy, nadciśnienia tętniczego oraz palenie tytoniu. Zmienna DEATH_EVENT określa, czy w okresie obserwacji nastąpił zgon pacjenta, i stanowi zmienną wynikową wykorzystywaną w procesie klasyfikacji. Charakterystykę poszczególnych atrybutów przedstawiono w tabeli 1.

Do analizy danych wykorzystano program WEKA w wersji 3.8.7 (Machine Learning Group, University of Waikato, New Zealand). Jest to oprogramowanie typu

open source przeznaczone do analizy danych i uczenia maszynowego, zawierające między innymi narzędzia do wstępnego przetwarzania danych, selekcji cech oraz klasyfikacji.

Do oceny jakości klasyfikacji zastosowano 10-krotną walidację krzyżową (10-fold cross-validation). W metodzie tej zbiór danych dzielony jest na 10 części o zbliżonej liczebności. W każdej iteracji dziewięć części wykorzystywanych jest do uczenia modelu, natomiast pozostała część służy do jego testowania. Procedura jest powtarzana dziesięciokrotnie, dzięki czemu każda część zbioru zostaje jednokrotnie wykorzystana jako zbiór testowy. Końcową ocenę modelu wyznacza się na podstawie wyników uzyskanych we wszystkich iteracjach.

Tab. 1. Charakterystyka atrybutów w zbiorze Heart Failure Clinical Records

Atrybuty	Objaśnienie	Jednostka
Wiek	Wiek pacjenta	lata
Anemia	Występowanie anemii	0 – nie, 1 – tak
Nadciśnienie tętnicze	Występowanie nadciśnienia tętniczego	0 – nie, 1 – tak
Fosfokinaza kreatynowa (CPK)	Aktywność fosfokinazy kreatynowej we krwi	μg/L
Cukrzyca	Występowanie cukrzycy	0 – nie, 1 – tak
Frakcja wyrzutowa	Odsetek krwi wypompowywanej z lewej komory podczas skurczu	%
Płeć	Płeć pacjenta	0 – kobieta, 1 – mężczyzna
Stężenie kreatyniny w surowicy	Stężenie kreatyniny w surowicy krwi	mg/dL
Liczba płytek krwi	Liczba płytek krwi	tys./μL
Stężenie sodu w surowicy	Stężenie sodu w surowicy krwi	mEq/L
Palenie tytoniu	Informacja, czy pacjent pali tytoń	0 – nie, 1 – tak
Czas obserwacji	Długość okresu obserwacji pacjenta	dni
Zgon pacjenta (DEATH_EVENT)	Informacja, czy pacjent zmarł w okresie obserwacji	0 – przeżycie, 1 – zgon

Analizę przeprowadzono dla dwóch wariantów zestawu danych. Model 1 obejmował pełny zestaw dostępnych atrybutów. W Modelu 2 zastosowano selekcję cech przeprowadzoną za pomocą metody InfoGainAttributeEval oraz algorytmu Ranker. Do dalszej analizy wybrano cztery najwyżej ocenione atrybuty: czas obserwacji, stężenie kreatyniny w surowicy, frakcję wyrzutową oraz wiek pacjenta. Zmienną wynikową w obu modelach stanowił zgon pacjenta (DEATH_EVENT).

Do przeprowadzenia klasyfikacji wykorzystano cztery algorytmy:

- Naive Bayes – probabilistyczny klasyfikator oparty na twierdzeniu Bayesa. Algorytm zakłada warunkową niezależność cech względem zmiennej decyzyjnej, co pozwala na stosunkowo proste i szybkie wyznaczenie prawdopodobieństwa przynależności obiektu do określonej klasy.
- J48 – algorytm tworzenia drzew decyzyjnych będący implementacją algorytmu C4.5. Klasyfikacja odbywa się poprzez kolejne podziały zbioru danych na podstawie wartości cech, prowadzące do utworzenia reguł decyzyjnych przypisujących analizowane przypadki do odpowiednich klas [Khanna, 2021].
- Random Forest – zespołowa metoda uczenia maszynowego wykorzystująca zbiór drzew decyzyjnych. Poszczególne drzewa budowane są na podstawie losowo wybranych próbek danych i podzbiorów cech, natomiast końcowy wynik klasyfikacji wyznaczany jest na podstawie wyników uzyskanych przez poszczególne drzewa [Wikipedia, 2026].
- Logistic – klasyfikator wykorzystujący regresję logistyczną do określenia prawdopodobieństwa przynależności obserwacji do określonej klasy. W implementacji dostępnej w programie WEKA stosowana jest regularyzacja typu ridge, ograniczająca ryzyko nadmiernego dopasowania modelu do danych uczących [Lee, 2025].

Do oceny i porównania skuteczności klasyfikatorów wykorzystano następujące miary: TP Rate, FP Rate, Precision, F-Measure oraz ROC Area.

- TP Rate (True Positive Rate) – określa odsetek przypadków należących do klasy pozytywnej, które zostały prawidłowo sklasyfikowane jako pozytywne.
- FP Rate (False Positive Rate) – określa odsetek przypadków należących do klasy negatywnej, które zostały błędnie sklasyfikowane jako pozytywne.
- Precision – określa udział prawidłowo sklasyfikowanych przypadków pozytywnych wśród wszystkich przypadków zaklasyfikowanych przez model jako pozytywne.
- F-Measure – średnia harmoniczna wartości Precision i Recall (TP Rate), uwzględniająca zarówno precyzję, jak i czułość klasyfikatora.
- ROC Area (AUC) – pole powierzchni pod krzywą ROC, określające zdolność modelu do rozróżniania analizowanych klas. Wartości bliższe 1 oznaczają lepszą zdolność klasyfikacyjną modelu.

3. Wyniki badań

3.1. Wyniki klasyfikacji dla pełnego zestawu atrybutów

Wyniki klasyfikacji uzyskane dla Modelu 1, obejmującego pełny zestaw 13 atrybutów, przedstawiono w tabeli 2. Spośród analizowanych algorytmów najlepsze wyniki uzyskał klasyfikator Random Forest. Wartość TP Rate wyniosła 0,849, Precision – 0,847, F-Measure – 0,847, natomiast ROC Area osiągnęła wartość 0,911. Jednocześnie Random Forest charakteryzował się najniższą wartością FP Rate, wynoszącą 0,219.

Dla klasyfikatora Logistic uzyskano TP Rate równy 0,829, Precision równą 0,826, F-Measure równą 0,826 oraz ROC Area równą 0,869. Klasyfikator J48 osiągnął TP Rate równy 0,806, Precision równą 0,803 oraz ROC Area równą 0,750. Najniższe wartości większości analizowanych miar uzyskano dla klasyfikatora Naive Bayes, dla którego TP Rate wyniósł 0,776, Precision – 0,770, F-Measure – 0,761, natomiast FP Rate osiągnął najwyższą wartość równą 0,380.

Tab. 2. Wyniki klasyfikacji dla pełnego zestawu atrybutów (Model 1)

Klasyfikatory	TP Rate	FP Rate	Precision	F-Measure	ROC Area
J48	0,806	0,273	0,803	0,804	0,750
Random Forest	0,849	0,219	0,847	0,847	0,911
Naive Bayes	0,776	0,380	0,770	0,761	0,852
Logistic	0,829	0,256	0,826	0,826	0,869

Źródło: opracowanie własne.

3.2. Wyniki klasyfikacji po selekcji cech

W Modelu 2 klasyfikację przeprowadzono z wykorzystaniem czterech atrybutów wybranych w procesie selekcji cech: czasu obserwacji, stężenia kreatyniny w surowicy, frakcji wyrzutowej oraz wieku pacjenta. Wyniki klasyfikacji przedstawiono w tabeli 3.

Po przeprowadzeniu selekcji cech wartości TP Rate dla klasyfikatorów J48, Naive Bayes i Logistic wzrosły odpowiednio z 0,806 do 0,823, z 0,776 do 0,806 oraz z 0,829 do 0,833. Jednocześnie wartości FP Rate dla tych klasyfikatorów zmniejszyły się odpowiednio z 0,273 do 0,260, z 0,380 do 0,344 oraz z 0,256 do 0,244.

Największą zmianę zaobserwowano dla klasyfikatora Naive Bayes, dla którego TP Rate wzrósł o 0,030, a Precision z 0,770 do 0,806. Wartość ROC Area wzrosła z 0,852 do 0,859. W przypadku J48 wartość ROC Area wzrosła z 0,750 do 0,785, natomiast dla klasyfikatora Logistic z 0,869 do 0,883.

Dla Random Forest po selekcji cech odnotowano niewielkie obniżenie wartości analizowanych miar. TP Rate zmniejszył się z 0,849 do 0,839, Precision z 0,847 do 0,837, F-Measure z 0,847 do 0,838, natomiast ROC Area z 0,911 do 0,909. Wartość FP Rate wzrosła z 0,219 do 0,230. Pomimo tych zmian Random Forest zachował najwyższą wartość ROC Area spośród analizowanych klasyfikatorów.

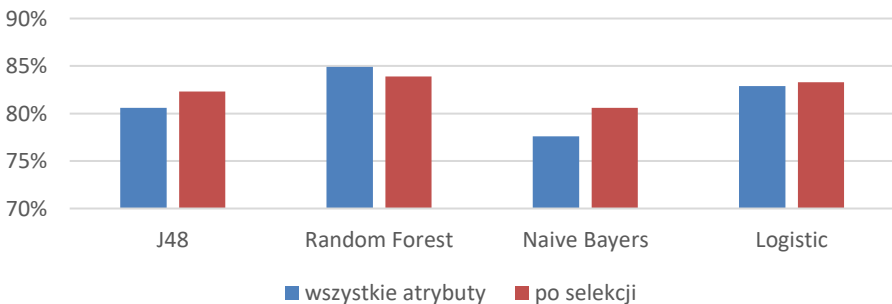
Tab. 3. Wyniki klasyfikacji po selekcji cech (Model 2)

Klasyfikatory	TP Rate	FP Rate	Precision	F-Measure	ROC Area
J48	0,823	0,260	0,819	0,820	0,785
Random Forest	0,839	0,230	0,837	0,838	0,909
Naive Bayes	0,806	0,344	0,806	0,793	0,859
Logistic	0,833	0,244	0,830	0,830	0,883

Źródło: opracowanie własne.

3.3. Porównanie modeli i analiza drzewa decyzyjnego

W celu porównania skuteczności analizowanych modeli zestawiono odsetek poprawnie sklasyfikowanych przypadków uzyskany przed i po przeprowadzeniu selekcji cech. Wyniki przedstawiono na rys. 1.

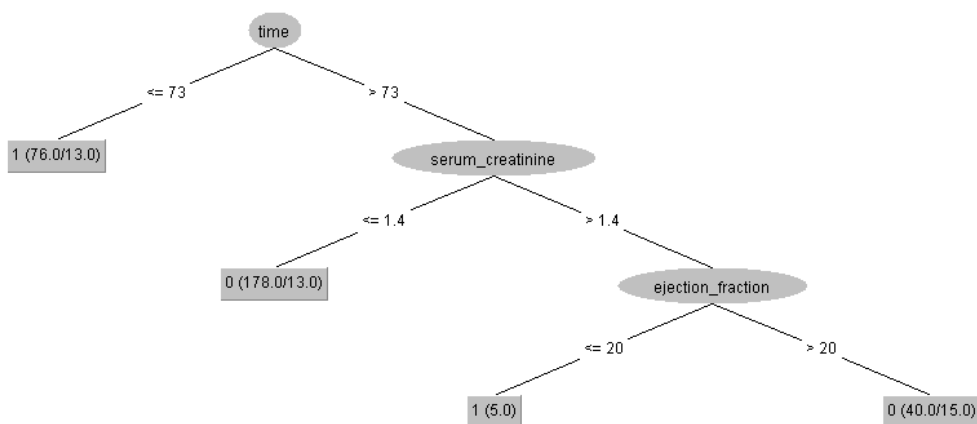


Rys. 1. Porównanie odsetka poprawnie sklasyfikowanych przypadków dla Modelu 1 i Modelu 2

Źródło: opracowanie własne.

Po przeprowadzeniu selekcji cech wzrost odsetka poprawnie sklasyfikowanych przypadków odnotowano dla klasyfikatorów J48, Naive Bayes oraz Logistic. Największą poprawę uzyskano dla klasyfikatora Naive Bayes, bo aż o 3 punkty procentowe. W przypadku Random Forest odnotowano niewielkie zmniejszenie odsetka poprawnie sklasyfikowanych przypadków. Dla wszystkich klasyfikatorów wykorzystujących zredukowany zestaw cech odsetek poprawnie sklasyfikowanych przypadków przekroczył 80%.

Dodatkowo dla klasyfikatora J48 utworzono drzewo decyzyjne umożliwiające przedstawienie sposobu klasyfikacji w postaci reguł decyzyjnych. Drzewo uzyskane dla Modelu 2 przedstawiono na rys. 2.



Rys. 2. Drzewo decyzyjne uzyskane za pomocą klasyfikatora J48 dla Modelu 2

Źródło: opracowanie własne.

Na podstawie utworzonego drzewa wyznaczono następujące reguły decyzyjne:

- jeżeli czas obserwacji ≤ 73 dni, wówczas decyzja = 1;
- jeżeli czas obserwacji > 73 dni oraz stężenie kreatyniny w surowicy $\leq 1,4$ mg/dL, wówczas decyzja = 0;
- jeżeli czas obserwacji > 73 dni, stężenie kreatyniny w surowicy $> 1,4$ mg/dL oraz frakcja wyrzutowa $\leq 20\%$, wówczas decyzja = 1;
- jeżeli czas obserwacji > 73 dni, stężenie kreatyniny w surowicy $> 1,4$ mg/dL oraz frakcja wyrzutowa $> 20\%$, wówczas decyzja = 0.

Wartość decyzji równa 1 oznacza zgon pacjenta w okresie obserwacji, natomiast wartość 0 oznacza przeżycie pacjenta.

4. Dyskusja wyników

Przeprowadzona analiza wykazała, że zastosowane metody uczenia maszynowego umożliwiają klasyfikację przeżywalności pacjentów z niewydolnością serca na podstawie dostępnych danych klinicznych. Spośród analizowanych algorytmów najlepsze wyniki dla pełnego zestawu atrybutów uzyskał Random Forest. Klasyfikator ten osiągnął najwyższe wartości TP Rate, Precision, F-Measure i ROC Area, a jednocześnie najniższą wartość FP Rate. Wskazuje to na jego wysoką skuteczność w analizowanym zadaniu klasyfikacyjnym.

Istotnym elementem przeprowadzonego badania była ocena wpływu selekcji cech na jakość klasyfikacji. Ograniczenie liczby atrybutów z 13 do 4 nie spowodowało wyraźnego pogorszenia skuteczności analizowanych modeli. W przypadku klasyfikatorów J48, Naive Bayes oraz Logistic odnotowano poprawę części analizowanych miar, natomiast dla Random Forest zaobserwowano jedynie niewielkie pogorszenie wyników. Oznacza to, że zastosowanie odpowiednio dobranego, ograniczonego zestawu cech może pozwolić na zachowanie dobrej jakości klasyfikacji przy jednoczesnym zmniejszeniu złożoności modelu.

W wyniku przeprowadzonej selekcji cech do Modelu 2 wybrano czas obserwacji, stężenie kreatyniny w surowicy, frakcję wyrzutową oraz wiek pacjenta. Szczegółowe znaczenie stężenia kreatyniny w surowicy oraz frakcji wyrzutowej jest zgodne z wynikami przedstawionymi przez Chicco i Jurmana [Chicco i Jurman, 2020]. Autorzy ci wykazali, że parametry te należą do najważniejszych czynników umożliwiających przewidywanie przeżywalności pacjentów z niewydolnością serca.

Znaczenie wymienionych parametrów znalazło również odzwierciedlenie w strukturze drzewa decyzyjnego uzyskanego za pomocą klasyfikatora J48. Poza czasem obserwacji w regułach decyzyjnych wykorzystane zostały stężenie kreatyniny w surowicy oraz frakcja wyrzutowa. W utworzonym modelu wartości stężenia kreatyniny przekraczające 1,4 mg/dL oraz frakcji wyrzutowej nieprzekraczające 20% były związane z klasyfikacją pacjenta do grupy, w której wystąpił zgon w okresie obserwacji. Wyniki te pozostają zgodne z kierunkiem zależności opisanym przez Chicco i Jurmana [Chicco i Jurman, 2020].

Przeprowadzona analiza wskazuje również, że skuteczność klasyfikacji zależy zarówno od zastosowanego algorytmu, jak i zestawu wykorzystywanych cech. Selekcja atrybutów może być zatem istotnym etapem przygotowania danych do budowy modeli predykcyjnych, szczególnie w przypadku zbiorów zawierających wiele parametrów klinicznych.

Należy jednocześnie uwzględnić ograniczenia przeprowadzonego badania. Analizowany zbiór obejmował 299 pacjentów, dlatego uzyskanych wyników nie należy

bezpośrednio uogólniać na całą populację pacjentów z niewydolnością serca. Wyniki wymagają dalszej weryfikacji na większych i niezależnych zbiorach danych. Modele uczenia maszynowego powinny być przy tym traktowane jako potencjalne narzędzia wspomagające ocenę rokowania, a nie jako rozwiązania zastępujące ocenę kliniczną lekarza.

Podsumowanie

Przeprowadzone badania wykazały możliwość zastosowania wybranych metod uczenia maszynowego do klasyfikacji przeżywalności pacjentów z niewydolnością serca na podstawie analizowanych danych klinicznych. Spośród zastosowanych algorytmów najlepsze wyniki dla pełnego zestawu atrybutów uzyskał klasyfikator Random Forest, osiągając najwyższą wartość ROC Area oraz korzystne wartości pozostałych miar jakości klasyfikacji.

Selekcja cech pozwoliła ograniczyć liczbę atrybutów z 13 do 4 przy zachowaniu zbliżonej skuteczności klasyfikacji. W przypadku algorytmów J48, Naive Bayes oraz Logistic zastosowanie zredukowanego zestawu cech przyniosło poprawę części analizowanych miar, natomiast dla Random Forest zaobserwowano jedynie niewielkie pogorszenie wyników.

Wśród cech wybranych w procesie selekcji znalazły się stężenie kreatyniny w surowicy oraz frakcja wyrzutowa, których znaczenie w prognozowaniu przeżywalności pacjentów z niewydolnością serca wskazano również w badaniach Chicco i Jurmana [Chicco i Jurman, 2020]. Uzyskane wyniki potwierdzają zasadność stosowania selekcji cech podczas budowy modeli klasyfikacyjnych oraz wskazują na potencjał metod uczenia maszynowego jako narzędzi wspomagających analizę danych klinicznych i ocenę rokowania pacjentów.

ORCID iD

Fabian Żebrowski: <https://orcid.org/0009-0009-6537-0305>

Małgorzata Zdrodowska: <https://orcid.org/0000-0003-4383-5713>

Literatura

1. Narodowy Instytut Kardiologii (b.r.), *Niewydolność serca*, <https://www.ikard.pl/niewydolnosc-serca.html>.
2. American Heart Association (2025), *Classes and Stages of Heart Failure*, <https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/classes-of-heart-failure>.
3. Narodowy Instytut Kardiologii (2025), *Narodowy Program Chorób Układu Krążenia na lata 2022–2032. Raport analityczny za 2024 rok*, https://bip.ikard.pl/images/dzialalnosc-instytutu/NPchUK/2024/Raport%20analityczny%20za%202024%20rok_KRMWK.pdf.
4. UCI Machine Learning Repository (2020), *Heart Failure Clinical Records*, Dataset, DOI: 10.24432/C5Z89R.
5. Chicco D., Jurman G. (2020), Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone, *BMC Medical Informatics and Decision Making* 20, 16, DOI: 10.1186/s12911-020-1023-5.
6. Khanna N. (2021), *J48 Classification (C4.5 Algorithm) in a Nutshell*, Medium, <https://medium.com/@nilimakhanna1/j48-classification-c4-5-algorithm-in-a-nutshell-24c50d20658e>.
7. Wikipedia (2026), *Las losowy*, https://pl.wikipedia.org/wiki/Las_losowy.
8. Lee F. (2025), *What Is Logistic Regression?*, IBM, <https://www.ibm.com/think/topics/logistic-regression>.
9. Brownlee J. (2019), *How to Perform Feature Selection With Machine Learning Data in Weka*, Machine Learning Mastery, <https://machinelearningmastery.com/perform-feature-selection-machine-learning-data-weka>.
10. Raphael C., Briscoe C., Davies J., Whinnett Z.I., Manisty C., Sutton R., Mayet J., Francis D.P. (2007), *Limitations of the New York Heart Association Functional Classification System and Self-Reported Walking Distances in Chronic Heart Failure*, *Heart* 93 (4), pp. 476-482.
11. Buchan T.A., Ross H.J., McDonald M., Billia F., Delgado D., Duero Posada J.G., Luk A., Guyatt G.H., Alba A.C. (2019), *Physician Prediction versus Model Predicted Prognosis in Ambulatory Patients with Heart Failure*, *Journal of Heart and Lung Transplantation* 38 (4), p. 381.

Predicting Survival in Patients with Heart Failure

Abstract

Heart failure is a major clinical problem associated with high mortality and a complex set of factors affecting patient prognosis. The aim of this study was to evaluate the applicability of machine learning methods for predicting survival in patients with heart failure and to assess the effect of feature selection on classification performance. The analysis was conducted using the Heart Failure Clinical Records dataset, which contains clinical data from 299 patients with left ventricular systolic dysfunction. Four classification algorithms—Naive Bayes, J48, Random Forest, and logistic regression—were evaluated using 10-fold cross-validation. Models based on the complete set of attributes were compared with models using features selected with the InfoGainAttributeEval method and the Ranker search algorithm. The best results for the complete dataset were obtained with the Random Forest classifier, which achieved a TP Rate of 0.849, Precision of 0.847, and ROC Area of 0.911. Feature selection improved classification performance for most of the evaluated algorithms while reducing the number of variables required by the models. The results indicate that machine learning methods can effectively support survival prediction in patients with heart failure and that appropriate feature selection can reduce model complexity without substantially compromising predictive performance.

Key words

heart failure, machine learning, patient survival, feature selection