Paweł Marciniak

Analiza i implementacja metod sztucznej inteligencji przy niepełnej informacji medycznej na przykładzie oceny ryzyka chorób

Łódź 2021

Recenzenci:

prof. dr hab. inż. Krzysztof Waldemar Wawryn

prof. dr hab. inż. Andrzej Handkiewicz

ISBN 978-83-66741-33-1

DOI: 10.34658/9788366741331

Wydawnictwo Politechniki Łódzkiej

93-005 Łódź, ul. Wólczańska 223

Tel. 42-631-20-87, 42-631-29-52

E-mail: zamowienia@info.p.lodz.pl

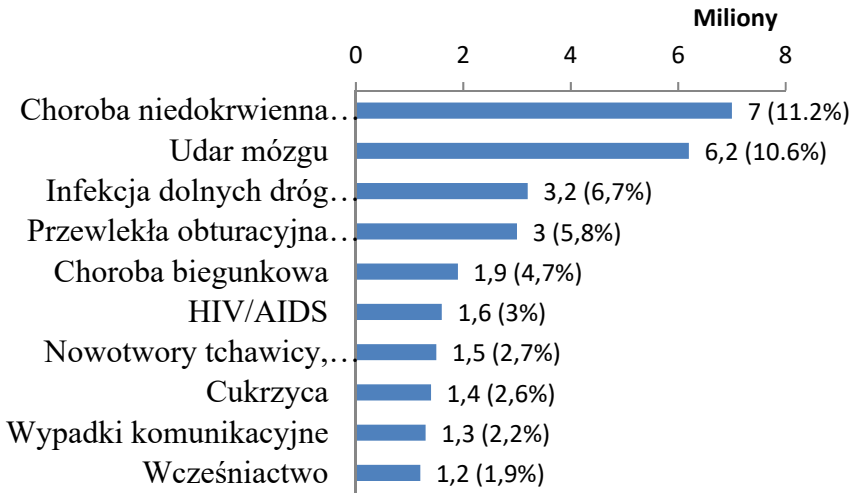
www.wydawnictwo.p.lodz.pl

Monografie Politechniki Łódzkiej, Nr 2391

SPIS TREŚCI

1. WPROWADZENIE

Choroby układu krążenia są najczęstszą przyczyną hospitalizacji (16% w grupie 6 400 000 pacjentów hospitalizowanych w 2006 roku) i stanowią największy odsetek zgonów (46% ogółu zgonów w 2006 roku) w Polsce [1, 2]. Dane dotyczące śmiertelności na świecie wyglądają podobnie. Badania przeprowadzone w 2011 roku przez Światową Organizację Zdrowia (ang. World Health Organization) również ukazują choroby serca jako główną przyczynę zgonów na świecie.



Rysunek 1.1. Przyczyny śmiertelności na świecie w 2011 roku według Światowej Organizacji Zdrowia

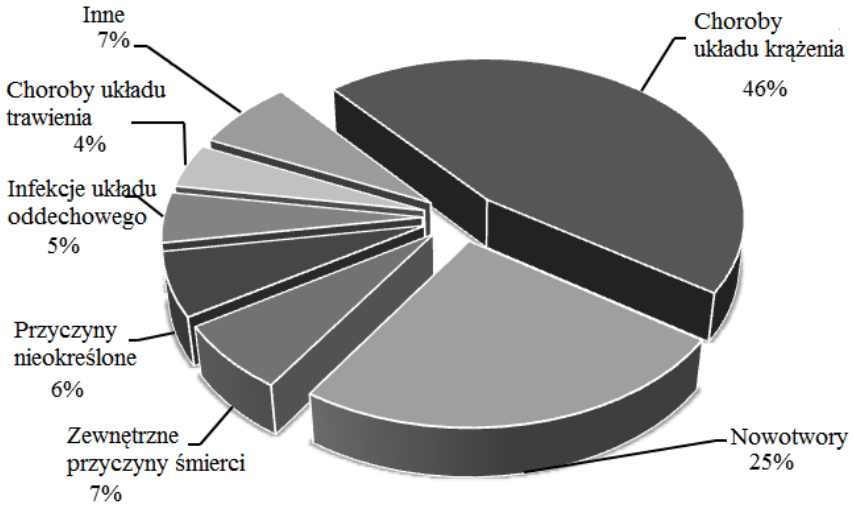
Źródło: opracowanie własne.

Bezpośrednią przyczyną śmiertelności szpitalnej jest w głównej mierze nagłe zatrzymanie krążenia (NZK). Choroby układu krążenia są znacznie częstszą przyczyną przedwczesnych zgonów wśród mieszkańców Polski niż przeciętnie w Unii Europejskiej i w ostatnich latach w Polsce spadek przedwczesnej umieralności z ich powodu uległ niestety spowolnieniu.

Pomimo że odsetek zgonów w Unii Europejskiej jest niższy niż w Polsce (około 40% wszystkich zgonów), szacuje się, że związane z tą grupą chorób obciążenie finansowe dla systemów opieki zdrowotnej UE wynosiło w 2006 roku nieco poniżej 110 mld euro. Oznacza to roczny koszt 223 euro na jed-

nego mieszkańca, czyli około 10 procent całkowitych wydatków na opiekę zdrowotną w UE. Choroby układu krążenia, jako jedne z najczęstszych chorób przewlekłych, mają negatywne skutki dla rynku pracy i są silnie powiązane z warunkami społecznymi.

Wprowadzenie standardów resuscytacji i postęp w technikach monitorowania funkcji życiowych nie rozwiązały wszystkich problemów związanych z ryzkiem nagłego zatrzymaniem krążenia u chorych leczonych w szpitalach. Według danych ze szpitali brytyjskich w wyniku resuscytacji udaje się przywrócić funkcje życiowe u 43% pacjentów z NZK, ale jedynie 30% przeżywa następne 24 godziny. Do czasu wypisania ze szpitala przeżywa jedynie 22% chorych [3].



Rysunek 1.2. Przyczyny śmierci w Polsce w 2006 roku

Źródło: [1].

Poziom zdrowia ludności analizuje się przede wszystkim na podstawie wyników badań medycznych. Ma on silny wpływ na wszystkie dziedziny naszego życia. Za jeden z najważniejszych aspektów mających na celu poprawienie poziomu zdrowia uważa się poprawną interpretację danych medycznych i zaproponowanie prawidłowej diagnozy. Sytuacje, w których lekarze muszą brać pod uwagę duże ilości danych znacznie utrudniają podejmowanie poprawnych decyzji medycznych. Potrzebne jest w związku z tym narzędzie, które będzie w stanie pomóc im w postawieniu właściwej

diagnozy. Do tego celu mogą zostać zastosowane sztuczne sieci neuronowe, które są częścią systemów sztucznej inteligencji.

Systemy wspomagania diagnostyki medycznej rozwijane są od kilkudziesięciu lat. Różne metody sztucznej inteligencji stosowane są w wielu aplikacjach służących zarówno do rozpoznawania, jak i poprawnej klasyfikacji sygnałów wejściowych. Mając na uwadze dynamiczny ich rozwój, a także dotychczasowe aplikacje tych metod do zagadnień medycznych, celowe wydaje się zastosowanie algorytmów sztucznej inteligencji do analizy ryzyka chorób kardiologicznych.

Biorąc pod uwagę wykazaną na wstępie potrzebę obiektywizacji i automatyzacji oceny stanu pacjenta, co wiąże się z możliwością zmniejszenia śmiertelności wewnątrzszpitalnej, uzasadnione wydają się badania nad algorytmami pozwalającymi na taką ocenę. Dają one bowiem szansę na zbliżenie standardu opieki nad pacjentem do postępowania optymalnego. Do najczęściej wykorzystywanych i najskuteczniejszych metod sztucznej inteligencji należą: sztuczne sieci neuronowe, sieć Bayesa, algorytmy genetyczne [4, 5], drzewa decyzyjne, logika rozmyta.

Znaczenie algorytmów klasyfikacji i predykcji można rozpatrywać w kontekście procesu leczenia pacjenta oraz w kontekście rozwoju wiedzy medycznej. Wykorzystanie algorytmów sztucznej inteligencji potrzebujących niewielkich mocy obliczeniowych (możliwych do zaimplementowania na prostym systemie wbudowanym) może przy stosunkowo niewielkich kosztach podnieść jakość opieki nad pacjentem, zmniejszyć ryzyko błędu, poprawić bezpieczeństwo pacjenta i personelu medycznego, a co najważniejsze obniżyć ryzyko nagłego zatrzymania krążenia, co może przekładać się na obniżenie śmiertelności wewnątrzszpitalnej.

Potwierdzeniem istotności tej tematyki jest przyznanie przez Narodowe Centrum Nauki finansowania projektu pt. „Zautomatyzowany system wieloparametrowej oceny stanu ogólnego pacjenta z pogłębioną analizą funkcji układu oddechowego i układu krążenia". Głównym celem tego projektu było badanie metodyki oceny ogólnego stanu chorego, ryzyka nagłego zatrzymania krążenia (NZK) oraz wspomagania decyzji dotyczącej dalszego leczenia wykorzystując sztuczne sieci neuronowe. Ich zastosowanie może zmniejszyć ryzyko NZK i śmiertelności szpitalnej. Wyżej wymienione algorytmy działają na podstawie danych zbieranych w czasie rzeczywistym oraz informacji wprowadzonych przez personel medyczny.

Projekt miał na celu przede wszystkim badanie zastosowania sztucznych sieci neuronowych oraz sieci Bayesa w medycynie ze szczególnym uwzględnieniem:

- zaawansowanej analizy zgromadzonych danych oceniających stan pacjenta,
- oceny ryzyka nagłego zatrzymania krążenia,
- sugerowanej metody dalszego leczenia,
- wskazania najbardziej prawdopodobnej przyczyny pogorszenia stanu zdrowia pacjenta.

Najistotniejszym punktem projektu było rozszerzenie aktualnego stanu wiedzy na temat adaptacji sztucznych sieci neuronowych oraz sieci Bayesa do oceny ogólnego stanu pacjenta i oceny ryzyka NZK. Dane wejściowe pobierane są w trybie rzeczywistym z monitora funkcji życiowych. Zadaniem sieci neuronowych oraz sieci Bayesa jest bardziej precyzyjna ocena stanu pacjenta na podstawie sygnałów z monitora funkcji, wykrycie sytuacji świadczących o ryzyku NZK, a w szczególności wykrycie cech niestabilności elektrycznej serca, ostrej niewydolności serca, niewydolności oddechowej, odwodnienia lub przewodnienia, ciężkiej infekcji oraz określenie sugestii co do dalszego postępowania.

Najprostszym podejściem jest porównywanie parametrów rejestrowanych przebiegów z pewnymi wartościami progowymi. Podejście takie, z uwagi na odgórnie założone wartości progów, nie uwzględnia jednak specyfiki przebiegu parametrów u poszczególnych pacjentów, nie umożliwia także zaawansowanego wnioskowania na podstawie grupy przesłanek – poszczególne parametry są zwykle rozpatrywane całkowicie oddzielnie, a co najwyżej łączone w sposób nieuwzględniający wagi poszczególnych parametrów, ich korelacji itp.

Ocena stanu ogólnego pacjenta wymaga czasu, wiedzy i doświadczenia, toteż wprowadzenie systemu wspomagania decyzji wydaje się znaczną pomocą pozwalającą na zaoszczędzenie czasu i obniżenie ryzyka błędu.

Niezbędne badania spirometryczne i echokardiograficzne prowadzone były przez członków zespołu na atestowanym sprzęcie dostępnym w Klinice Pneumonologii i Alergologii z Pododdziałem Kardiologicznym Szpitala Klinicznego Nr 1 Uniwersytetu Medycznego w Łodzi. Do akwizycji danych i testów wykorzystany jest zakupiony przez Katedrę Mikroelektroniki i Technik Informatycznych kardiomonitor. Zakupiony monitor wyprodukowany został przez polską firmę Emtel, która zgodziła się na współpracę w ramach prowadzonych prac poprzez udostępnienie niezbędnych protokołów, jak

i wsparcie techniczne przy obsłudze sprzętu. KMiTI uzyskała zgodę komisji bioetycznej przy Uniwersytecie Medycznym w Łodzi na wykonanie badań klinicznych.

Praktyczne zastosowanie urządzenia jako uzupełnienie monitorów funkcji życiowych w jednostkach medycznych daje nadzieję na lepszą kwalifikację do hospitalizacji i zmniejszenie ryzyka popełnienia błędu w ocenie ogólnego stanu pacjenta. Ponadto pozwoli na wczesne wykrycie zagrożenia wystąpienia NZK i w efekcie zmniejszenie śmiertelności szpitalnej. Ważnym efektem będzie też wyznaczenie parametrów krzywej oddechowej mogących mieć znaczenie diagnostyczne i prognostyczne, określenie przedziałów ich wartości normalnych, a także przedstawienie możliwości zastosowania parametrów uzyskanych z krzywej oddechowej i sygnału EKG zamiast kardiografii impedancyjnej.

Głównym celem badań było opracowanie efektywnych algorytmów metod sztucznej inteligencji dla nowatorskiego systemu monitorującego, wspomagającego w czasie rzeczywistym podejmowanie decyzji medycznych, związanych z oceną stanu ogólnego pacjenta ze szczególnym uwzględnieniem parametrów kardiologicznych. Danymi wejściowymi do systemu (wykonanego w postaci zminiaturyzowanego urządzenia) są wielorakie parametry zbierane przez typowy monitor funkcji życiowych.

Podstawowym zadaniem postawionym przed opisanymi badaniami jest analiza parametrów fizjologicznych pacjenta pozwalająca na zminimalizowanie zbioru danych wejściowych dla algorytmów sztucznej inteligencji przy zachowaniu pierwotnej jakości klasyfikacji oraz zwiększenie jakości klasyfikacji przy zastosowaniu pierwotnej liczby parametrów. Obydwa zagadnienia są istotne z punktu widzenia monitorowania stanu pacjenta w trybie rzeczywistym i możliwe są do zaimplementowania na systemie mikroprocesorowym niewielkiej mocy dołączonym do monitora funkcji życiowych.

Niniejsza monografia powstała na bazie doktoratu pod tytułem „Analiza i implementacja metod sztucznej inteligencji przy niepełnej informacji medycznej na przykładzie oceny ryzyka chorób kardiologicznych".

1.1. Obecnie stosowane systemy do oceny stanu ogólnego pacjenta

1.1.1. Modified Early Warning Score (MEWS)

Zmodyfikowany Wskaźnik Wczesnego Ostrzegania (MEWS) to prosty współczynnik używany do szybkości określenia stanu pacjenta. Jest on oparty na danych pochodzących z czterech odczytów parametrów fizjologicznych (skurczowe ciśnienie krwi, tętno, częstość oddechów, temperatura ciała) i jedną obserwację na podstawie wpisu pielęgniarki – poziom świadomości AVPU:

- Alert (świadomy),
- Voice (reaguje na głos),
- Pain (reaguje na ból),
- Unresponsive (brak reakcji).

Wyniki każdego z parametrów należy ocenić w skali od 0 do 3 zgodnie z tabelą 1.1 [6, 7].

Tabela 1.1. Współczynnik MEWS

Obserwacje	**Wynik**						
	3	2	1	0	1	2	3
Skurczowe ciśnienie krwi	<45%	30%	15% mniej	Naturalne dla pacjenta	15% więcej	30%	>45%
Puls [uderzenia/min.]	-	<40	41-50	51-100	101-110	111-129	>130
Częstość oddechów [oddechy/min]	-	<9	-	9-14	15-20	21-29	>30
Temperatura [°C]	-	<35	-	35.0-38.4	-	>38.5	-
AVPU	-	-	-	A	V	P	U

Źródło: opracowanie własne.

1.1.2. APACHE III

Acute Physiology, Age and Chronic Health Evaluation (APACHE III) jest wskaźnikiem opierającym się na metodzie wprowadzonej w 1991 roku przez Knausa, składającym się z punktów [7]:

- za nieprawidłowości fizjologiczne (zakres od 0 do 252) – 17 zmiennych fizjologicznych odzwierciedlających wartości życiowe, badania laboratoryjne i stan neurologiczny;
- za wiek (od 0 do 24);
- za współistniejące choroby przewlekłe (od 0 do 23) [8].

Sumując wyniki z trzech wymienionych kategorii otrzymujemy ostateczną wartość współczynnika APACHE III (zakres od 0 do 299). Współczynnik ten może być używany do pomiaru stopnia zaawansowania choroby i porównywania stanu pacjentów w ramach jednej kategorii diagnostycznej lub niezależnie od posiadanego schorzenia.

1.1.3. Therapeutic Intervention Scoring System

Terapeutyczna Skala Interwencji Medycznych (TISS – The Therapeutic Intervention Scoring System) została wprowadzona w 1974 roku i stała się powszechnie stosowaną metodą klasyfikacji pacjentów znajdujących się w stanie krytycznym. Ze względu na ciągły rozwój dziedzin nauki stosowanych w oddziałach intensywnej opieki medycznej (medycyny, elektroniki, informatyki) wskaźnik ten był wielokrotnie aktualizowany.

W przeciwieństwie do innych systemów punktacji, TISS nie jest wskaźnikiem służącym do określenia stanu ogólnego pacjenta lub stopnia zaawansowania choroby. Skala TISS służy do określenia kosztów opieki nad pacjentem hospitalizowanym na Oddziale Intensywnej Opieki Medycznej oraz do kalkulacji czasu pracy personelu medycznego niezbędnego do pełnej opieki nad chorym. Jedna z odmian systemu TISS – TISS-28 – stosowana jest w Polsce. Formularz TISS-28 składa się z 28 pytań (odpowiedź tak/nie) podzielonych na 7 grup: czynności podstawowe, oddychanie, krążenie krwi, nerki, metabolizm, ośrodkowy układ nerwowy, inne interwencje [9].

1.2. Metodyka badań

W celu opracowania i porównania różnych podejść opartych o algorytmy z dziedzin sztucznej inteligencji, uczenia maszynowego i metod redukcji wymiarowości danych zaprojektowana została metodyka badań składająca się z następujących etapów:

1. Opracowanie uniwersalnej, modularnej aplikacji komputerowej, pozwalającej na przeprowadzanie zaawansowanych analiz, transformacji danych oraz trenowanie i testowanie opracowanych algorytmów sztucznej inteligencji;
2. Przygotowanie odpowiednich zestawów danych, które służyły zarówno do nauki, jak i testowania opracowanych algorytmów;
3. Opracowanie algorytmów transformacji danych wejściowych;
4. Zaprojektowanie i porównanie różnych algorytmów redukcji wymiarowości danych;
5. Adaptacja przygotowanych danych medycznych;
6. Redukcja parametrów medycznych;
7. Zaprojektowanie i porównanie algorytmów sztucznej inteligencji.

W ramach ostatniego z etapów autor przeprowadził badania bazujące na następujących podejściach:

- sztucznych sieciach neuronowych;
- sieci Bayesa;
- drzewach decyzyjnych;
- liniowym modelu decyzyjnym.

1.2.1. Sztuczne sieci neuronowe

Wymienienie wszystkich dziedzin nauki i życia, w jakich stosowane są sztucznie sieci neuronowe jest zadaniem trudnym. Ciągły rozwój metod sztucznej inteligencji jest motorem do prób zastosowania sieci neuronowych w miejscach, w których do tej pory nie były stosowane. Sztuczne sieci neuronowe wykorzystywane są w różnych dziedzinach, takich jak [10]:

- ekonomia:

  - prognozy giełdowe;
  - prognozy cen;
  - prognozy sprzedaży;
  - analiza problemów produkcyjnych;
  - optymalizacja działalności handlowej;
- statystyka;
- medycyna:
  - analiza badań medycznych;
  - badania psychiatryczne;
  - interpretacja badań biologicznych;
- przewidywanie wyniku zawodów sportowych;
- prognozowanie postępów w nauce;
- diagnostyka układów elektronicznych;
- selekcja celów śledztwa w kryminalistyce;
- analiza spektralna;
- planowanie remontów maszyn;
- poszukiwanie ropy naftowej;
- sterowanie procesów przemysłowych;
- klasyfikacja i identyfikacja obrazów, dokumentów, danych.

Zaletą sztucznych sieci neuronowych jest ich uniwersalność – mogą być zaimplementowane zarówno programowo, jak i w warstwie sprzętowej (np. w postaci układów VLSI [11, 12])

Algorytmy sztucznych sieci neuronowych mogą wypracowywać wnioski dzięki zakodowanej w nich wiedzy lekarskiej dotyczącej związków między elementami różnorodnych sytuacji fizjologicznych i patologicznych oraz umożliwić uczenie się systemu przez porównanie zestawu danych wejściowych oraz informacji pochodzących od lekarza (rozpoznanie).

Zastosowanie sztucznych sieci neuronowych w medycynie jest bardzo szerokie. SSN są stosowane m.in. przy:

- analizie przebiegów sygnałów EKG, EEG i innych – w szczególności przy wykrywaniu niebezpiecznych symptomów;
- analizie obrazów – rozpoznawaniu i interpretacji różnego typu obrazów z aparatury medycznej;
- klasyfikacji i ocenie stanu pacjenta – rozpoznawaniu symptomów oznaczających pogorszenie stanu pacjenta;

– wspomaganiu dalszego leczenia – optymalizacji sposobu leczenia, sugerowaniu odpowiedniej terapii [10].

1.3. Sieci bayesowskie

W XVIII wieku brytyjski matematyk i duchowny Thomas Bayes sformułował twierdzenie teorii prawdopodobieństwa, które pośmiertnie zostało nazwane twierdzeniem Bayesa. Omawiana reguła łączy ze sobą prawdopodobieństwo warunkowe zdarzeń P(X|Y) oraz P(Y|X). Zasada ta mówi nam, jak bezwarunkowe i warunkowe prawdopodobieństwa są ze sobą powiązane, czy mamy do czynienia z prawdopodobieństwem obiektywnym czy subiektywnym. Na przykład, jeśli X jest zdarzeniem „u pacjenta występuje wysoka gorączka", a Y jest zdarzeniem „pacjent ma grypę", twierdzenie Bayesa pozwala przeliczyć znany odsetek gorączkujących wśród chorych na grypę i znane odsetki gorączkujących i chorych na grypę w całej populacji, na prawdopodobieństwo, że ktoś jest chory na grypę, gdy wiemy, że ma wysoką gorączkę.

Sieci bayesowskie, będące zaawansowaną formą probabilistycznego podejścia opartego o rozumowanie bayesowskie, znalazły wiele zastosowań w aplikacjach medycznych [12, 13, 14]. Sieć bayesowska składa się z grafu skierowanego opisującego zależności jakościowe pomiędzy zdarzeniami oraz specyfikacji liczbowej opisującej zależności ilościowe.

Opis graficzny stosowany w sieciach bayesowskich jest przejrzysty, ekspresywny, łatwy do analizy i modyfikacji, pozwala w wygodny sposób zakodować proces decyzyjny naśladujący ludzkie rozumowanie. Graf pozwala określić związki przyczynowo-skutkowe między zdarzeniami, a tym samym pokazać, które zdarzenia są od siebie warunkowo niezależne. Dzięki koncepcji warunkowej niezależności sieć ma możliwość prowadzenia lokalnych obliczeń, obejmujących niewielki wycinek przestrzeni zmiennych, co zapewnia efektywność obliczeniową. Specyfikacja liczbowa jest podawana dla węzłów grafu i obejmuje prawdopodobieństwa bezwarunkowe i warunkowe, uzależnione od stanu węzłów-rodziców danego węzła.

W przeciwieństwie do np. sieci neuronowych, pełna specyfikacja (struktura, wartości liczbowe) jest zrozumiała i łatwa do analizy przez człowieka, może być też przez człowieka opracowana. Daje to możliwość zakodowania fragmentu wiedzy i doświadczenia lekarza w postaci algorytmu

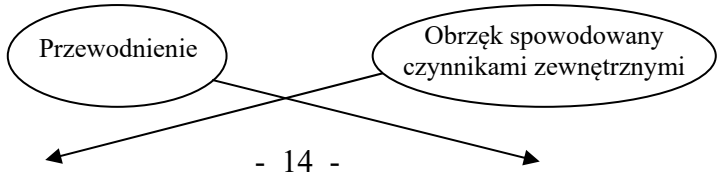
numerycznego. Sieci tego typu mają możliwość uczenia maszynowego, zarówno struktury, jak i specyfikacji liczbowej, na podstawie danych historycznych, co umożliwia ich opracowanie na podstawie badań klinicznych przeprowadzonych w projekcie. Dodatkowo, możliwe jest, w przypadku długotrwałego monitorowania pacjenta, dostosowanie algorytmu do jego specyfiki na podstawie danych pomiarowych i decyzji (rozpoznania) lekarza.

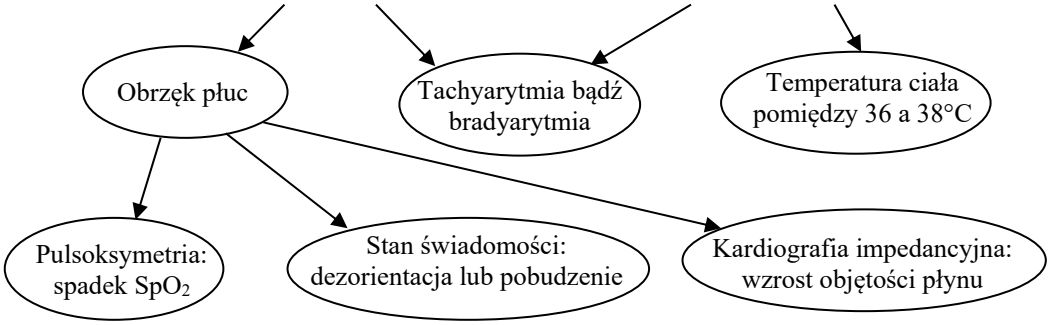
Analiza ryzyka chorób kardiologicznych w oparciu o sieci bayesowskie zostanie zilustrowana prostym przykładem. Na rys. 1.3 przedstawiono uproszczony fragment sieci wiążący ze sobą przesłanki sugerujące wystąpienie obrzęku płuc (węzły na dole rysunku), dwie przyczyny wystąpienia obrzęku (węzły na górze rysunku) oraz węzeł pozwalający na dyskryminację pomiędzy tymi przyczynami (dwa węzły po prawej stronie środkowego wiersza). Wszystkie węzły są dwustanowe. W tablicach prawdopodobieństw na potrzeby przykładu przyjęto identyczną wagę wszystkich przesłanek. Tak skonstruowaną sieć można wykorzystać do prowadzenia obliczeń; przykładowe wyniki są zebrane w tabeli 1.2. Jak widać, nawet tak prosta sieć potrafi uzależnić prawdopodobieństwo obrzęku płuc od przesłanek oraz rozróżniać pomiędzy jego przyczynami.

Tabela 1.2. Prawdopodobieństwa dla sieci z rysunku 1.3

Wejścia					Wyjścia [prawdopodobieństwo]	
Pulsoksymetria spadek SpO_2	Stan świadomości: dezorientacja lub pobudzenie	Kardiografia impedancyjna wzrost objętości	Tachyarytmia bądź arytmia	Temperatura ciała pomiędzy 36 a 38°C	Przewodnienie	Obrzęk spowodowany czynnikami zewnętrznymi
nie	nie	nie	nie	tak	0,016	0,006
tak	nie	nie	nie	tak	0,144	0,052
tak	tak	tak	nie	tak	0,925	0,335
tak	tak	tak	tak	tak	0,506	0,988

Źródło: opracowanie własne.





Rysunek 1.3. Sieć bayesowska

Źródło: opracowanie własne.

Sieci bayesowskie nie są jednak pozbawione wad – opierają się na reprezentacji w postaci atrybut-wartość, a ich struktura oraz specyfikacja numeryczna jest stała podczas procesu wnioskowania. Powoduje to np. niemożność opisu sytuacji, gdy dany węzeł może mieć zmienną liczbę rodziców albo sytuacji, w których prawdopodobieństwo danego stanu jest określone zależnością funkcyjną. Prace zmierzające do usunięcia tych ograniczeń doprowadziły do rozwiązania znanego jako Multi-Entity Bayesian Networks (MEBN) [15].

Sieci MEBN stanowią rozszerzenie sieci bayesowskich o rachunek predykatów pierwszego rzędu. W przypadku tych sieci opis składa się nie z pojedynczej sieci o określonej strukturze, a z kolekcji fragmentów – fragmenty takie określane są jako MFrag. Każdy MFrag przypomina swą strukturą typową sieć bayesowską. W stosunku do klasycznej sieci dodane są tu tzw. węzły kontekstowe, określające warunki, które muszą być spełnione, aby dany MFrag został włączony do procesu wnioskowania. Pozostałe węzły są podzielone na wejściowe i wyjściowe; MFrag definiuje rozkład prawdopodobieństwa węzłów wyjściowych warunkowany stanami węzłów wejściowych i węzłów kontekstowych. Wszystkie węzły mogą posiadać argumenty (zmienne) – dzięki temu jeden MFrag może opisywać związki pomiędzy dowolną liczbą konkretnych obiektów.

Zasadniczą różnicę widać w przypadku specyfikacji liczbowej. W rzeczywistości nie jest to już specyfikacja jedynie liczbowa – prawdopodobieństwa stanów węzłów wyjściowych są określone poprzez wyrażenia logiczne i funkcyjne, których argumentami są liczność i stany węzłów wejściowych. Takie podejście daje bardzo dużą elastyczność w definiowaniu specyfikacji.

Opis konkretnego problemu jest dokonywany poprzez kolekcję elementów MFrag. Niektóre elementy MFrag są „wbudowane" i dotyczą wszystkich problemów (np. MFrag opisujący pojęcie równości), inne są tworzone specjalnie na potrzeby opisu konkretnego problemu. Kolekcja elementów MFrag spełniająca warunek istnienia unikalnego łącznego rozkładu prawdopodobieństwa zmiennych losowych występujących we wszystkich elementach MFrag nosi nazwę MTheory.

1.4. Drzewa decyzyjne

Jedną z bardzo popularnych technik wykorzystywanych do analizy danych są drzewa decyzyjne [16]. Drzewem decyzyjnym nazywamy drzewograf, które można przedstawić jako graficzną metodę wspomagania procesu podejmowania decyzji (klasyfikacji). Obraz drzewa decyzyjnego składa się z:

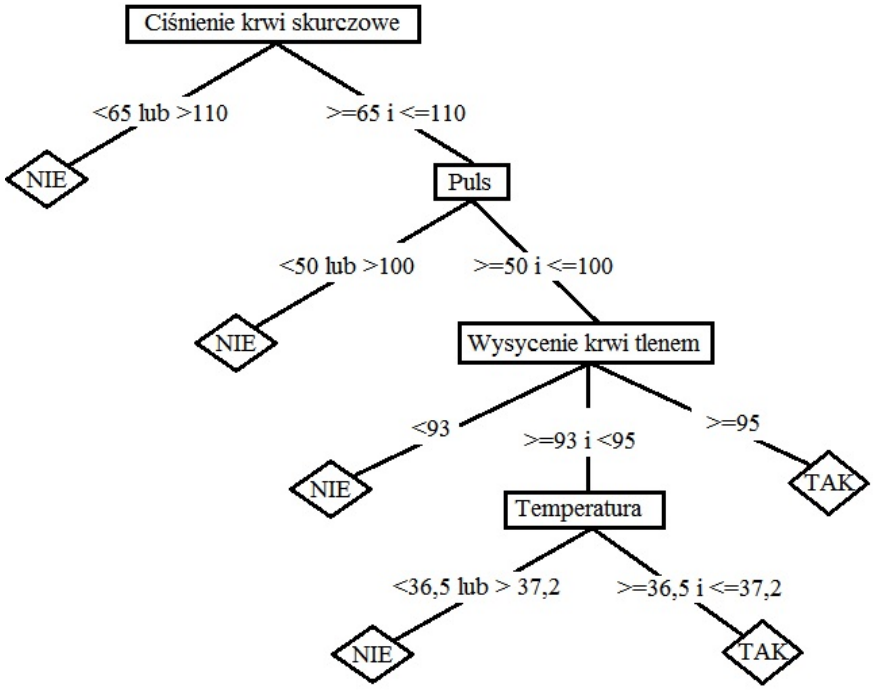
- korzenia (wierzchołka, w którym podejmowana jest pierwsza decyzja);
- gałęzi (łączących ze sobą wierzchołki, określających liczbę możliwych wariantów – wartości atrybutów);
- węzłów (wierzchołków, z których wychodzi co najmniej jedna gałąź);
- liści (wierzchołków końcowych – bez gałęzi wychodzących).

Drzewo decyzyjne jest grafem składającym się z wierzchołków (atrybutów) i gałęzi reprezentujących wartości tego atrybutu. Metoda drzewa decyzyjnego opiera się na analizie rożnych przypadków opisanych przez pewne zestawy atrybutów. Każdy z atrybutów przyjmuje pewne wartości dyskretne (w przypadku zmiennych ciągłych wskazana jest dyskretyzacja poprzez wprowadzenie przedziałów).

Odpowiednia klasyfikacja polega na przejściu od korzenia do liścia i przypisaniu do danego przypadku klasy (decyzji) zapisanej w liściu. Możliwe jest to dzięki podziale złożonego problemu na zadania, kolejno na mniejsze zadania i podzadania, aż do momentu, w którym będzie możliwe podjęcie jednoznacznej decyzji. W każdym wierzchołku sprawdzany jest pewien warunek (atrybut) dotyczący danego przypadku, i na jego podstawie wybierana jest jedna z gałęzi prowadząca do kolejnego węzła (liścia).

Na rysunku 1.4 przedstawiono przykładowe drzewo decyzyjne wspomagające wypisy pacjentów ze szpitala. Korzystając z tak skonstruowanego

drzewa decyzyjnego lekarz jest w stanie podjąć decyzję zgodną z przykładem zawartym w tabeli 1.3.



Rysunek 1.4. Drzewo decyzyjne

Źródło: opracowanie własne.

Budowa drzewa decyzyjnego na podstawie zbioru przykładów polega na rozpatrzeniu kolejnych wierzchołków (zaczynając od korzenia) i podjęciu decyzji, czy ten punkt drzewa ma być:

- liściem drzewa – taka decyzja spowoduje utworzenie węzła końcowego i zakończenie procesu klasyfikacji według kryterium stopu;
- węzłem – decyzja ta spowoduje wybór odpowiedniego testu według kryterium wyboru testu. Kolejno, na podstawie wartości testu, zostanie utworzona kolejna „korona drzewa" składająca się z tylu wierzchołków, ile jest wartości określających wybrany test. Dla węzłów z następnego

poziomu drzewa należy dokonać podobnego wyboru pamiętając o zmniejszeniu zestawu dostępnych testów o ostatnio wybrany.

Tabela 1.3. Przykładowe dane dla drzewa decyzyjnego

Wejścia						**Wyjście**
Wiek	**Płeć**	**Temperatura**	**Puls**	**Skurczowe ciśnienie krwi**	**Wysycenie krwi tlenem**	**Sugerowany koniec hospitalizacji**
65	M	35,2	63	43	96	Nie
32	K	38,3	41	108	92	Nie
79	K	37,4	70	81	98	Tak
44	M	36,6	98	72	94	Tak

Źródło: opracowanie własne.

Przedstawiony sposób budowy drzewa stosowany jest we wszystkich algorytmach konstruowania drzew decyzyjnych. Różnice między poszczególnymi algorytmami dotyczą przede wszystkim implementacji dwóch decyzji:

- o wyborze rodzaju wierzchołka (węzeł lub liść) – kryterium stopu;
- o wyborze testu dla węzła – kryterium wyboru testu.

Kryterium stopu określa, kiedy ma być zatrzymany proces rozrostu drzewa, czyli kiedy dla pozostałego zbioru przykładów wierzchołkiem może być jedynie liść. Z takim przypadkiem mamy do czynienia, gdy zbiór pozostałych obiektów:

- jest pusty;
- zawiera obiekty należące tylko do jednej klasy decyzyjnej – w takim przypadku kryterium powinno zwrócić tą klasę;
- nie ulega podziałowi przez żadne kryterium wyboru;
- zbiór dostępnych kryteriów wyboru jest pusty.

Kryterium wyboru testu jest bardzo istotnym elementem budowania drzewa decyzyjnego. To ono decyduje o złożoności drzewa. Jedną z najważniejszych zasad, którą należy się kierować przy wyborze testu dla węzłów jest prostota drzewa. W zależności od typów atrybutów mogą być używane różne rodzaje testów:

- testy tożsamościowe – test utożsamiany jest z atrybutem:

$$t(x) = a(x); \tag{1.1}$$

- testy równościowe – test sprawdzający równość:

$$t(x) = \begin{cases} 1 \; gdy \; a(x) = z \\ 0 \; gdy \; a(x) \neq z \end{cases}; \tag{1.2}$$

- testy przynależnościowe – test sprawdzający przynależność do zbioru:

$$t(x) = \begin{cases} 1 \; gdy \; a(x) \in Z \\ 0 \; gdy \; a(x) \notin Z \end{cases}; \tag{1.3}$$

- testy podziałowe – test sprawdzający przynależność do podzbiorów powstałych przez podział przeciwdziedziny atrybutu:

$$t(x) = \begin{cases} 1 \; gdy \; a(x) \in \; Z_1 \\ 2 \; gdy \; a(x) \in \; Z_2 \\ \quad \ldots \\ n \; gdy \; a(x) \in \; Z_n \end{cases}; \tag{1.4}$$

Bardzo często do oceny testu wykorzystuje się funkcje mierzące różnicę między zbiorem P (przykładów) a zbiorami, na jakie dzieli się ten zbiór według wartości ocenianego atrybutu ze względu na rozkład częstotliwości klas (decyzji). Przykładem takiego rodzaju oceny jest funkcja przyrostu informacji opierająca się na zastosowaniu entropii do pomiaru nierównomierności rozkładu kategorii. Funkcja przyrostu informacji wyrażona jest wzorem:

$$g_t(P) = I(P) - E_t(P), \tag{1.5}$$

gdzie $E_t(P)$ jest entropią zbiorów przykładów P:

$$E_{tr}(P) = \sum_{d \in C} -\frac{\left|P_{tr}^{d}\right|}{\left|P_{tr}\right|} \log \frac{\left|P_{tr}^{d}\right|}{\left|P_{tr}\right|}, \tag{1.6}$$

a I(P) jest informacja zawartą w zbiorze etykietowanych przykładów *P*.

$$I(P) = \sum_{d \in C} -\frac{\left|P^{d}\right|}{\left|P\right|} \log \frac{\left|P^{d}\right|}{\left|P\right|}. \tag{1.7}$$

Natomiast entropia zbioru przykładów P ze względu na atrybut *t* jest definiowana jako średnia ważona entropii dla poszczególnych podzbiorów tego zbioru:

$$E_t(P) = \sum_{r \in R_t} \frac{\left|P_{tr}\right|}{\left|P\right|} E_{tr}(P) \quad . \tag{1.8}$$

Wartość ta osiąga duże wartości w przypadku, gdy wartość atrybutu *t* dzieli przykładu ze zbioru t na równomiernie liczne podzbiory.

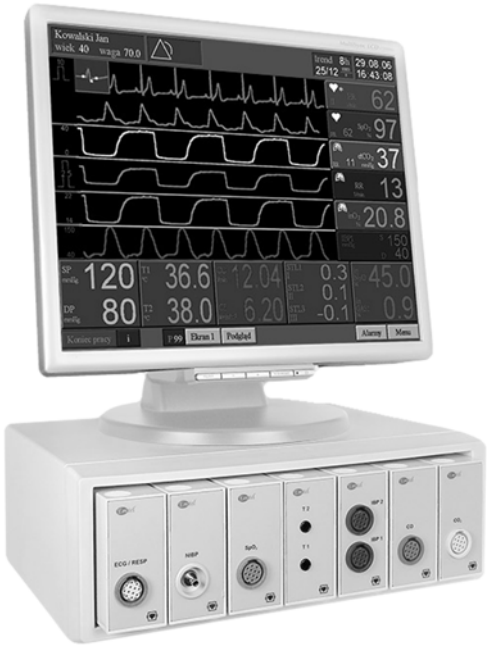
Niekiedy przy budowie drzewa decyzyjnego istotnym kryterium wyboru testu, poza jakością, może być koszt testu. Można sobie wyobrazić sytuację, w której dziedziną jest zbiór pacjentów w szpitalu, a atrybutami są wyniki różnego typu badań diagnostycznych. Wartości atrybutów będą dostępne dopiero po przeprowadzeniu takich badań. Koszt badania możemy rozpatrywać zarówno w odniesieniu do pieniędzy, jak i czasu potrzebnego na jego przeprowadzenie. Należy również pamiętać, że niektóre badania mogą rodzić niebezpieczeństwo utraty zdrowia przez pacjenta (to również jest koszt testu). Przy wyborze testu trzeba mieć na uwadze zarówno jakość, jak i koszt testu.

2. CZĘŚĆ BADAWCZA

Badania prowadzone w ramach niniejszej pracy stanowiły jedną z części interdyscyplinarnego projektu finansowanego przez Narodowe Centrum Nauki pod tytułem „Zautomatyzowany system wieloparametrowej oceny stanu ogólnego pacjenta z pogłębioną analizą funkcji układu oddechowego i układu krążenia", realizowanego wspólnie przez Katedrę Mikroelektroniki i Technik Informatycznych Politechniki Łódzkiej oraz Uniwersytecki Szpital Kliniczny nr 1 im. N. Barlickiego Uniwersytetu Medycznego w Łodzi.

Główną ideą projektu było opracowanie nowatorskiego systemu monitorującego, wspomagającego w czasie rzeczywistym podejmowanie decyzji medycznych, związanych z oceną stanu ogólnego pacjenta, na podstawie wielorakich danych zbieranych przez typowy monitor funkcji życiowych; system zostanie wykonany w postaci zminiaturyzowanego urządzenia dołączonego do wybranego modelu monitora. Istotnym z punktu widzenia medycyny jest również przeprowadzenie dogłębnej analizy krzywej oddechowej, wyznaczanej przez monitor funkcji życiowych, w celu określenia jej przydatności jako istotnego składnika oceny stanu pacjenta. W omawianym projekcie biorą udział pracownicy i doktoranci Katedry Mikroelektroniki i Technik Informatycznych Politechniki Łódzkiej oraz Uniwersytecki Szpital Kliniczny nr 1 im. N. Barlickiego Uniwersytetu Medycznego w Łodzi.

W ramach prac prowadzonych w projekcie autor brał czynny udział przy projektowaniu i realizacji aplikacji napisanej w języku C++, służącej do zbierania danych klinicznych z monitora funkcji życiowych. Transmisja parametrów odbywa się w trybie rzeczywistym. Za jej pomocą zapisywane są dane pacjentów z komercyjnego monitora funkcji życiowych. Jest to podstawą do implementacji i weryfikacji algorytmów sztucznej inteligencji.

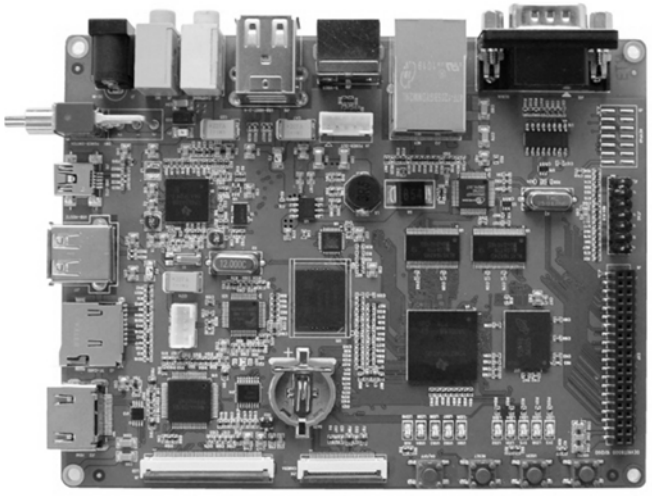


Rysunek 2.1. Monitor funkcji życiowych Emtel FX2000-MD

Źródło: [17].

Ze względu na fakt, że system sztucznej inteligencji ma działać w trybie rzeczywistym i powinien umożliwiać interakcję z personelem medycznym, został dokonany wybór sprzętu elektronicznego, na którym mają zostać zaimplementowane algorytmy decyzyjne.

Najprostszym wyborem spełniającym takie wymagania jest komputer przenośny wyposażony w ekran dotykowy. Ten wybór ma swoje wady – najważniejszą z nich jest obudowa takiej jednostki: dwa elementy połączone zawiasowo i zewnętrzne źródło energii nie gwarantują wymaganej solidności, zwłaszcza, jeśli urządzenie ma być stosowane w ambulansach. Innym możliwym rozwiązaniem byłby tablet lub smartfon z dużym ekranem. Problem w tym wypadku jest podobny, jak w przypadku standardowego komputera przenośnego i dodatkowo może wymagać zewnętrznych urządzeń peryferyjnych, takich jak zasilacz, głośnik, ewentualnie kluczkonwerter interfejsu Ethernet. Problemy te spowodowały, że do implementacji algorytmów wybrana została platforma Embest DevKit8500 z procesorem Cortex- A8, procesorem graficznym TMS320C64x oraz akceleratorem grafiki POWERVR SGX.



Rysunek 2.2. Platforma implementacji algorytmów Embest DevKit8500D

Źródło: [17].

Na wymienionej platformie sprzętowej mogą zostać zainstalowane i wspierane systemy operacyjne: Android, Ångström, Windows i Linux.

2.1. Aplikacja testowania algorytmów

W celu przetestowania omawianego modelu sztucznych sieci neuronowych autorzy zaprojektowali dedykowaną aplikację. Program ten został napisany w języku C# i posiada właściwości pozwalające na:

- modelowanie wielowarstwowych sieci neuronowych bez sprzężenia zwrotnego;
- wybór liczby warstw ukrytych;
- określenie metody nauczania (dwie nadzorowane metody nauczania zostały wdrożone – algorytm wstecznej propagacji błędów oraz jego modyfikacja w postaci metody gradientów sprzężonych;
- zdefiniowanie liczby neuronów w każdej warstwie ukrytej;
- ustawienie współczynnika uczenia;

- wybór metody wstecznej propagacji z współczynnikiem momentum. Wprowadzanie tego współczynnika pomaga uniknąć problemów oscylacji (które mogą wystąpić w podstawowej wersji algorytmu wstecznej propagacji błędów) i utknięcia w lokalnym minimum;
- zdefiniowanie współczynnika nachylenia funkcji aktywacji;
- dodanie wejścia Bias do warstwy wejściowej i do każdej z warstw ukrytych;
- zdefiniowanie warunku końca nauki. Oprogramowanie pozwala na zakończenie uczenia przez ustawienie maksymalnego błędu (który jest liczony po prezentacji wszystkich wzorców ze zbioru uczącego) lub określenie maksymalnej liczby iteracji;
- określenie początkowych wag. Użytkownik ma dwie możliwości: losowanie wag (w zakresie od -0,2 do 0,2) lub załadowanie wag z pliku;
- zapisanie wag (początkowych i końcowych) do pliku;
- normalizowanie zestawu danych (wejścia i wyjścia). Normalizacja nie dotyczy wejść dwustanowych (binarnych). Użytkownik może wybrać jeden z trzech sposobów normalizacji:

1) „binarny" – tryb ten klasyfikuje wejścia numeryczne (np. stężenia cholesterolu, ciśnienie spoczynkowe krwi) do pewnych grup, jak pokazano w tabeli 2.1. Przynależność do grupy jest zakodowana wartościami binarnymi (nowymi wejściami sieci). Korzystając z tej metody liczba wejść zwiększa się (3-4 grupy – 2 wejścia, 5-8 grupy – 3 wejścia, itp.).

Tabela 2.1. Normalizacja „binarna"

Ciśnienie spoczynkowe krwi (CSK)		
Wejście [x_1]	Warunek	Wejścia [x_1, x_2] po normalizacji
80	CSK<100	0,0
120	100≤CSK<135	0,1
160	135≤CSK<170	1,0
220	170≤CSK<205	1,1

Źródło: opracowanie własne.

2) „binarny rozszerzony" – tryb ten klasyfikuje wejścia numeryczne (takie jak poziom cholesterolu, ciśnienie spoczynkowe krwi) do

pewnych grup, jak pokazano w tabeli 2.2. W tym przypadku liczba wejść (po normalizacji) jest zawsze równa liczbie grup.

Tabela 2.2. Normalizacja „binarna rozszerzona"

Ciśnienie spoczynkowe krwi (CSK)		
Wejście [x_1]	Warunek	Wejścia [x_1, x_2] po normalizacji
80	CSK<100	0,0,0,1
120	100≤CSK<135	0,0,1,0
160	135≤CSK<170	0,1,0,0
220	170≤CSK<205	1,0,0,0

Źródło: opracowanie własne.

3) „min–max" – normalizacja ta może być opisana za pomocą następującego równania:

$$f(x) = \frac{x - minx(x)}{\max(x) - \min(x)}, \qquad (2.1)$$

gdzie min(x) i max(x) są to odpowiednio minimalne i maksymalne wartości ze zbioru wartości danego wejścia. W tym przypadku liczba wejść (przed i po normalizacji) jest stała.
W wyniku użycia normalizacji „binarnej" i „binarnej rozszerzonej" można uzyskać na wejściu tylko dwie wartości – {0,1}. Użycie normalizacji „min-max" powoduje, że na wejściu sieci może się pojawić dowolna liczba z przedziału [0,1].

Badania skuteczności metod proponowanych w niniejszej pracy zostały przeprowadzone na podstawie danych z dwóch baz danych.

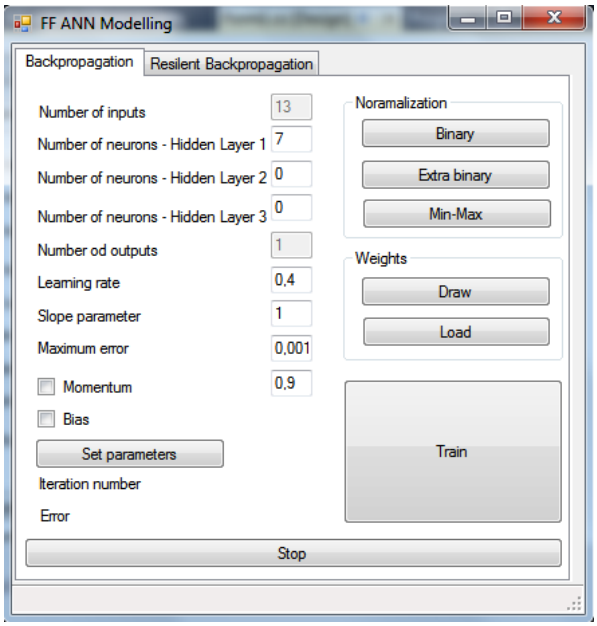
Baza danych **Cleveland** została opracowana przez dr. Robert Detrano z V.A. Medical Center, Long Beach and Cleveland Clinic Foundation. Baza ta składa się z danych zapisanych podczas badania 303 pacjentów. Sześciu pacjentów z powodu błędów z zapisie niektórych parametrów zostało odrzuconych. Każdy pacjent opisany został przez 13 parametrów medycznych (pola wejściowe). Czternasty parametr (wyjściowy) oznacza ogólny stan pacjenta ze szczególnym uwzględnieniem chorób serca. Parametr ten przyjmuje wartości od 0 do 4, gdzie wartość 0 oznacza zdrowego pacjenta, natomiast wartość 4 dotyczy pacjenta w najgorszym stanie. Opisy wejść i przyjmowanych przez wejścia wartości przedstawione są w tabeli 2.3.

Tabela 2.3. Opis parametrów w bazie danych Cleveland

Parametr i opis	Opis przyjmowanych wartości
Wiek	wartość numeryczna;
Płeć	0 – kobieta; 1 – mężczyzna.
Cp – typ bólu w klatce piersiowej	1 – typical angina; 2 – atypical angina; 3 – non-anginal pain; 4 – asymptomatic.
Trestbps – ciśnienie skurczowe w chwili przyjęcia do szpitala [mmHg]	wartość numeryczna.
Chol - serum cholesterol [mg/dl]	wartość numeryczna.
Fbs – zawartość cukru we krwi na czczo (>120 mg/dl)	0 – nie; 1 – tak.
Restecg – elektrokardiografia	0 – normalna; 1 – anormalność odcinka ST-T; 2 – możliwy lub zdiagnozowany przerost lewej komory.
Thalach – maksymalny zmierzony puls	wartość numeryczna.
Exang – ćwiczenia fizyczne powodują duszności	0 – nie; 1 – tak.
Oldpeak – obniżenie odcinka ST podczas ćwiczeń	wartość numeryczna.
Slope – nachylenie odcinka ST	1 – rosnący; 2 – płaski; 3 – malejący.
Ca - liczba dużych naczyń krwionośnych pokolorowanych podczas fluoroskopii	0-3.
Thal - koronarografia	3 – normalna; 6 – zmiany odwracalne; 7 – zmiany nieodwracalne.

Źródło: opracowanie własne.

Druga baza danych **Łódź** zebrana została przez lekarzy jednego z łódzkich szpitali klinicznych. Zawiera dane od 73 pacjentów opisanych przez 67 wielorakich parametrów pochodzących z różnego typu badań fizjologicznych. Dotyczą one między innymi sygnału EKG, parametrów chemicznych krwi, ogólnego stanu i wydolności serca. W bazie tej znajdują się również informacje na temat pacjenta spisywane podczas przyjęcia do szpitala – są to np. wiek, płeć. Parametr wyjściowy wpisany do bazy odwołuje się do wywiadu zrobionego z pacjentem po okresie dwóch lat od hospitalizacji.



Rysunek 2.3. Program do modelowania sztucznych sieci neuronowych

Źródło: opracowanie własne.

2.2. Transformacje danych

Najnowsze metody analizy danych, nazywane metodami eksploracyjnymi (data mining), polegają nie tylko na estymacji parametrów i sprawdzaniu hipotez, ale dotyczą redukcji wymiarowości, klasyfikacji i modelowania złożonych procesów, zjawisk, zachowań. Naukowcy z wielu specjalności (biolodzy, chemicy, informatycy, socjologowie) rozwijają tę

dziedzinę nauki w celu uzyskiwania bardziej wiarygodnych i dokładniejszych modeli statystycznych. Weryfikacja takiego modelu odbywa się poprzez wykazanie i udowodnienie, że posiada on zadowalający (zgodny z przyjętym założeniem) stopień dokładności w naturalnym środowisku badania.

Z uwagi na ograniczenia wielu procedur statystycznych, w praktyce stosuje się wstępne przygotowanie danych. Jest to bardzo istotny etap w projektowaniu modelu i można go podzielić na trzy części: transformację danych, uzupełnienie danych niepełnych i zmniejszenie liczby danych wejściowych (wymiarowości).

W praktyce, przy analizie różnego rodzaju danych doświadczalnych, często mamy do czynienia z występowaniem różnych skal pomiarowych. Dotyczy to zarówno sytuacji, w których występują cechy o różnych charakterze (np. ilościowe, jakościowe) oraz poszczególne cechy określone są przez różne wielkości fizyczne, mające różne zakresy wartości. W celu ujednolicenia wpływu poszczególnych cech można wykonać pewne przekształcenia (dla każdej cechy transformację należy przeprowadzić z osobna). Do najbardziej znanych można zaliczyć:

- **normalizację** – przeprowadzana jest według wzoru:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{2.2}$$

  gdzie x_{max} jest maksymalną wartością występującą w zbiorze treningowym dla i-tej cechy, x_{min} – minimalną wartością dla *i*-tej cechy. Przy wykonaniu takiego przekształcenia zarówno dla zbioru zbioru treningowego i testowego używane są same wartości x_{max} oraz x_{min}. Zgodnie z tą formułą otrzymuje się wektory, których wartości cech są z zakresu [0,1]. Ten sposób transformacji nie uwzględnia rozkładu wartości danej cechy. Oznacza to, że w przypadku wystąpienia obserwacji znacznie odstających od wartości typowych, następuje ściśnięcie dużej liczby informacji w wąskim przedziale.
- **standaryzację** – przeprowadzana jest według równania:

$$x' = \frac{x - \bar{x}}{\sigma}, \tag{2.3}$$

  gdzie, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ oraz $\sigma = \frac{1}{n-1}\sum_{i=1}^{n}(a_i - \bar{a}_i)$.
  Przekształcenie to powoduje otrzymanie cech, których wartość średnia wynosi 0, a odchylenie standardowe 1. Oznacza to, że wszystkie

zmienne mają takie same znaczenie przy liczeniu odległości. Wykorzystując metodę standaryzacji należy zachować ostrożność, gdyż użycie jej dla wektora cech, którego odchylenie standardowe jest bliskie zeru, może wprowadzić do danych duży szum. Drugim ograniczeniem jest przypadek, gdy $\sigma = 0$ – otrzymana zostanie wielkość nieokreślona (dzielenie przez zero). Z tego względu przed wykonaniem standaryzacji należy usunąć tzw. „płaskie wzorce" (*flat pattern*).

- **transformację nieliniową** – asymetria danych najczęściej nie jest pożądana – oznacza odstępstwo od rozkładu normalnego wymaganego przez wiele procedur statystycznych. W większości przypadków udaje się jej pozbyć (dotyczy to również sytuacji tzw. długich „ogonów", czyli gdy zadania rzadkie sumarycznie dominują w zbiorze) poprzez nieliniową transformację danych.

  Można w takim przypadku zastosować transformację rozkładów asymetrycznych prawostronnie ($\frac{1}{x^2}, \frac{1}{x}, \log x, \ln x, \sqrt{x}$) lub lewostronnie ($x^2, x^3$). Dla rozkładów asymetrycznych lewostronnie najsilniejszym przekształceniem jest $\frac{1}{x^2}$, a najsłabszym $\sqrt{x}$. W niektórych transformatach nie można zastosować wartości ujemnych – w takim przypadku należy przesunąć dane tak, aby wszystkie były dodatnie.

2.3. Redukcja wymiarowości danych

W rozdziale tym zaprezentowane zostały matematyczne narzędzia służące redukcji wymiaru danych w klasyfikacji. Przedstawione metody stosowane są w eksploracyjnej analizie danych z wielowymiarowej przestrzeni. Zadaniem tych rozwiązań jest wyrażenie wielowymiarowych obserwacji przy użyciu małej liczby współrzędnych, przy zachowaniu oryginalnych relacji między nimi. Odbywa się to poprzez identyfikację i eliminację nieistotnych nadmiarowych informacji w zbiorze danych. Redukcja wymiarowości jest etapem przygotowującym dane, który może być wdrożony przed właściwym algorytmem klasyfikacji, regresji lub innych metod sztucznej inteligencji. Cechy otrzymane w wyniku tego etapu mogą być następnie wykorzystane przez inne procedury analiz wielowymiarowych. Poprzez zastosowanie tego typu analiz możliwe jest zwizualizowanie danych wielowymiarowych na płaszczyźnie lub wykresie trójwymiarowym. Niekiedy do wyznaczenia istot-

nych cech można wykorzystać wykres współrzędnych równoległych (ang. *parallel coordinates plot*). Prezentuje on zmienne wielowymiarowe w postaci wykresu dwuwymiarowego, pozwala również określić, jak zmieniają się wartości poszczególnych cech dla tych samych obserwacji [18].

Redukcja danych pozwala odnieść następujące korzyści:

- zwiększyć efektywność uczenia – redukcja wymiaru umożliwia wybór istotnych cech z punktu widzenia zadania klasyfikacji. Pozwala to zredukować efekt przeuczenia, polegający na zbyt mocnym dopasowaniu klasyfikatora do danych uczących;
- zmniejszyć czas uczenia – w niektórych metodach wzrost liczby wymiarów powoduje więcej niż proporcjonalny wzrost czasu działania algorytmu;
- zmniejszyć wymagania dotyczące maszyny obliczeniowej;
- zmniejszyć złożoność struktury danych wejściowych – umożliwia to lepszą interpretację zidentyfikowanych sygnałów;
- wyeliminować błędy i szumy ze zbioru uczącego – pozwala to na eliminację cech nieistotnych dla zadania klasyfikacji;
- uzyskać kompromis pomiędzy poziomem kompresji danych wejściowych a skutecznością działania klasyfikatora.

Matematyczne metody mające na celu redukcję danych mogą być oparte o **konstrukcję cech**. W takim modelu n cech wejściowych zostaje poddanych pewnej transformacji, w wyniku której nowe zmienne wyjaśniają całość (lub zdecydowaną większość) zakodowanej informacji przy zmniejszonej liczbie danych wejściowych. Przykładami takiego typu analizy są:

- analiza czynnikowa;
- analiza składowych głównych;
- skalowanie wielowymiarowe.

Inną matematyczną metodą mającą na celu redukcję wymiaru jest **selekcja cech**. Polega ona na wyborze pewnego podzbioru spośród zbioru oryginalnych cech. Rozwiązanie to w odniesieniu do klasyfikacji polega na wyborze cech najlepiej dyskryminujących populację. Najważniejszym zadaniem selekcji cech na potrzeby klasyfikacji jest zwiększenie dokładności algorytmu klasyfikującego. Możliwość taka istnieje w przypadku, gdy nie wszystkie cechy niosą istotną informację (z punktu widzenia klasyfikacji).

Może zdarzyć się tak, że dana zmienna w przypadku jednego rodzaju klasyfikacji pozbawiona jest jakiejkolwiek wartości użytecznej (nie jest związana z rozpatrywanym w danym momencie badaniem), a w przypadku innej klasyfikacji niesie ze sobą bardzo istotną informację.

Jeżeli jedna z cech dla wszystkich rodzajów klasyfikacji (badań) nie zawiera istotnych informacji, może to oznaczać, że zmienna ta jest źle określona lub błędnie zmierzona. Duża liczba cech zaszumionych często może negatywnie wpływać na zdolności dyskryminacyjne niektórych modeli. W takim przypadku zmniejszenie liczby danych wejściowych może pozytywnie wpłynąć na zdolności predykcyjne zastosowanych metod sztucznej inteligencji. Jeżeli zmniejszenie liczby cech wejściowych nie powoduje zmniejszenia skuteczności dyskryminacji, to można zauważyć, że odejmowane cechy są zaszumione lub są to cechy istotne, ale nie wnoszące dodatkowej informacji.

Biorąc pod uwagę sposób relacji metody selekcji z innymi algorytmami nadrzędnymi można wyróżnić:

- **Metody rankingowe (filtry)** – klasa metod, która wyboru cech istotnych dla dalszego procesu uczenia dokonuje w sposób autonomiczny. Decyzję podejmuje niezależny od klasyfikatora współczynnik („informacja wzajemna", „dywergencja Kullbacka-Leiblera"). Metody rankingowe nie uwzględniają zależności między zmiennymi. Podejście to znajduje uzasadnienie w przypadku, gdy rozpatrywane zmienne są niezależne. Jeżeli w danym zbiorze danych występują korelacje pomiędzy danymi, zastosowanie tego rodzaju metod może okazać się nieefektywne.
- **Metody opakowane (wrappery)** – klasa metod używających do wyboru istotnych zmiennych informacji zwrotnej pomiędzy elementem decyzyjnym (siecią neuronową) a algorytmem selekcji. Metody te dokonują selekcji pod kątem konkretnego klasyfikatora. Miarą jakości wyselekcjonowanego podzbioru jest dokładność klasyfikatora.
- **Frappery** – kombinacja metod opakowanych i metod rankingowych. W tego rodzaju metodach do selekcji właściwych cech wykorzystywana jest metoda rankingowa – jednak parametry filtrów dostrajane są na podstawie metody opakowującej.
- **Metody wbudowane** – klasa metod, która używa pewnych cech algorytmów uczenia dokonując automatycznej selekcji cech podczas uczenia algorytmu decyzyjnego.

2.3.1. Informacja wzajemna

Jedną z popularnych metod wyboru istotnych cech jest obliczenie, jak dobrze każda ze zmiennych objaśnia zmienną decyzyjną [19]. Do tego celu można zastosować współczynnik informacji wzajemnej, oparty na teorii informacji Shannona. Do obliczenia współczynnika informacji wzajemnej dwóch danych pochodzących ze zbioru A oraz B można posłużyć się wzorem:

$$I(A,B) = \sum_x \sum_y p(x,y) \log_2 \frac{p(x,y)}{p_1(x)p_2(y)}, \qquad (2.4)$$

gdzie A i B są zbiorami, w których znajdują się zmienne (cechy) x i y – odpowiednio $A=\{x_1, x_2, \ldots, x_m\}$ oraz $B=\{y_1, y_2, \ldots, y_n\}$. Prawdopodobieństwo $p_1(x)$ wyraża szansę otrzymania danej x ze zbioru A, natomiast prawdopodobieństwo $p_2(x)$ oznacza szansę otrzymania cechy y ze zbioru B. Wyrażenie $p(x,y)$ oznacza prawdopodobieństwo otrzymania pary komunikatów (x,y) odpowiednio ze zbiorów A i B.

Współczynnik informacji wzajemnej mierzy jak bardzo znajomość cechy A ułatwia (statystycznie) przewidywanie wartości B. Jest to parametr nieujemny i osiąga wartość 0 jedynie w przypadku, gdy dane ze zbiorów A i B występują niezależnie od siebie, tzn. $p(x,y) = p_1(x)p_2(y)$. Można to interpretować tak, że jest to sytuacja, w której znajomość jednej cechy nie daje żadnej informacji na temat drugiej cechy.

Współczynnik informacji wzajemnej można również wyrazić poprzez entropię. W takim przypadku wyrażony on jest wzorem:

$$I(A,B) = H(A) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y), \qquad (2.5)$$

gdzie $H(A)$ oraz $H(B)$ oznaczają entropie (średnią ilość informacji w zmiennej losowej), $H(A,B)$ jest entropią łączną dla dwóch cech losowych A, B, natomiast $H(A|B)$ oznacza entropię warunkową dla pary zmiennych losowych A i B (ile dodatkowej informacji niesie ze sobą zmienna A, jeżeli znamy zmienną B).

Entropia, w teorii informacji, definiowana jest jako średnia ilość informacji, przypadająca na pojedynczą wiadomość ze źródła informacji. Można to interpretować jako średnią ważoną ilości informacji niesionej przez

pojedynczą wiadomość w przypadku, gdy wagami są prawdopodobieństwa wysłania poszczególnych wiadomości.

2.3.2. Analiza składowych głównych

Analiza czynników głównych (*Principal Components Analysis* – PCA) jest jedną z najpopularniejszych narzędzi do analizy danych i oceny występujących w danych zależności. Metoda ta używana jest do zmniejszenia wymiarowości danych przy zachowaniu możliwie największej ilości informacji. Technika ta jest nieparametryczna (nie wymaga żadnych założeń co do rozkładów badanych danych) oraz, w przeciwieństwie do większości metod selekcji cech, jest uczona bez nadzoru (nie jest uwzględniana przynależność wektorów danych do poszczególnych klas). Analiza składowych głównych może być wykorzystana również do znajdowania korelacji pomiędzy danymi binarnymi [20, 21, 22].

Podstawowym celem tej metody jest zastąpienie wejściowego wektora danych skorelowanych (w przypadku, gdy dane nie są skorelowane metoda nie zapewnia możliwości redukcji danych bez utraty informacji) poprzez mniejszą liczbę zmiennych nieskorelowanych, które potrafią wyjaśnić całą (lub prawie całą) zmienność danych oryginalnych. Odbywa się to poprzez znalezienie w przestrzeni danych kierunków o pewnych określonych własnościach. Kierunki mogą być interpretowane jako szukane dane nieskorelowane i określa się je mianem składowych głównych. Składowe te są kombinacjami liniowymi cech wejściowych i wybierane są tak, aby wyjaśniały jak najwięcej zmienności (w pierwszej składowej zakodowane jest najwięcej informacji, kolejne składowe nie są skorelowane z poprzednimi i zawierają coraz mniej informacji). Dzięki temu otrzymujemy w nowej przestrzeni składowe, które wyjaśniają całą zmienność oryginalnych danych.

W procesie odwzorowania przestrzeni danych wejściowych w nową przestrzeń składowych głównych nie zachodzi żadne skalowanie. Nowa przestrzeń odwzorowuje zarówno odległości, jak i kierunki (kąty) [23, 24].

Metoda PCA zwraca kilka rodzajów współczynników opisujących każdą składową główną:

- **Wartości własne** – informują, jaka część całkowitej zmienności tłumaczona jest przez daną składową. Z uwagi na fakt, że kolejne składowe

tłumaczą coraz mniej zmienności, wartości własne kolejnych składowych układają się w ciąg nierosnący. Wariancja całkowita jest sumą wartości własnych wszystkich składowych głównych. Pozwala to na wyliczenie procentu zmienności określonego przez daną składową, skumulowanej zmienności oraz skumulowanego procentu zmienności.

- **Wektor własny** – współczynnik odzwierciedlający wpływ poszczególnych zmiennych oryginalnych na daną składową główną.
- **Ładunki** – parametry obrazujące wkład poszczególnych zmiennych oryginalnych w składowe główne. Są to wartości określające, jaką część wariancji danej składowej stanowią zmienne oryginalne.
- **Wkłady zmiennych** – określają, jaki procent zmienności danej składowej głównej może być tłumaczony zmiennością poszczególnych zmiennych oryginalnych.

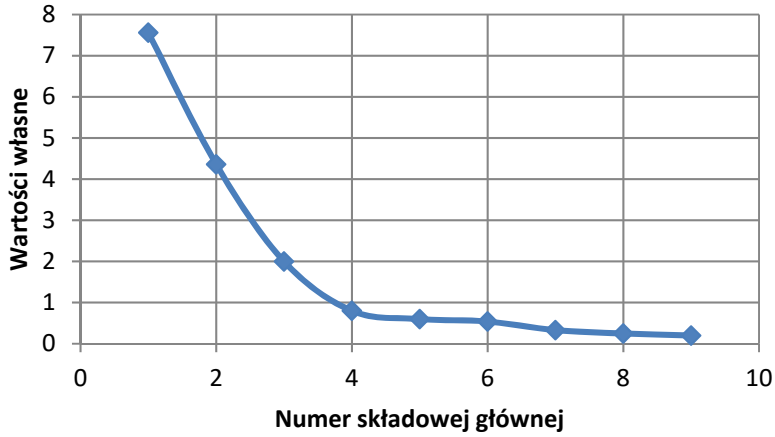
Bardzo istotnym elementem przy wykorzystaniu analizy składowych głównych do redukcji wymiarowości jest kryterium wyboru odpowiedniej liczby składowych. Nie istnieje jedno uniwersalne, najlepsze dla każdego przypadku kryterium wyboru liczby nowych czynników. Jako podstawowy cel należy uznać taką redukcję, przy której przy jednoczesnym pozbyciu się danych „nieistotnych", stanowiących niepożądany przy dalszej analizie szum, udaje się zachować jak najwięcej informacji istotnej z punktu widzenia późniejszej klasyfikacji.

Wartości własne, odpowiadające składowym głównym, są parametrem pomocnym, wyjaśniającym stopień powiązania składowej z całością informacji niesionej przez dane, ale nie dają jednoznacznej odpowiedzi, ile składowych należy wybrać do dalszej analizy. Przez lata stosowania metody składowych głównych powstało kilka, często stosowanych, praktycznych kryteriów pomocnych przy podjęciu decyzji. Poniżej przedstawiono najpopularniejsze:

- ***Kryterium Kaisera-Guttmana*** – mówi ono, że składowe główne, które chcemy zachować, powinny posiadać wartość wariancji niemniejszą niż dowolna wystandaryzowana zmienna oryginalna. Odpowiada to wartości własnej składowej większej od jedności. Argumentem przemawiającym za korzystaniem z tego kryterium są wątpliwości przy stosowaniu zmiennych o mniejszym wkładzie wariancji niż wkład pojedynczej zmiennej pierwotnej. Zaletą tego kryterium jest możliwość zautomatyzowania wyboru liczby składowych

głównych. Kryterium to nie wymaga wizualnej interpretacji wykresu wartości własnych (wykres osypiska).

- **Kryterium stopnia wyjaśnionej wariancji**. Liczba składowych głównych, które należy zostawić do dalszej analizy określana jest na podstawie stopnia wyjaśnienia, który chcemy uzyskać. Wszystkie składowe główne zgodnie z założeniem niosą 100% wariancji zmiennych pierwotnych. Stosując to kryterium należy z góry założyć pewien procent wariancji, jaki chcemy uzyskać i wybrać do realizacji tego założenia odpowiednią liczbę składowych. Najczęściej przyjmuje się, że składowe główne powinny wyjaśniać 80-90% wariancji. Tak jak w przypadku kryterium Kaisera możliwa jest prosta automatyzacja procesu wyboru liczby składowych.
- **Analiza równoległa** – w metodzie tej generowane jest n obserwacji z rozkładu $N(0,1)$ dla każdej zmiennej z osobna, następnie obliczana jest macierz korelacji, dla której poszukiwana jest wartość własna (krok ten powtarzany jest k razy). Kolejno obliczana jest wybrana statystyka pozycyjna (kwantyl) dla każdej zmiennej. Liczba składowych określona jest przez liczbę wartości własnych większych od kwantyla (oraz większych od 1).
- **Kryterium osypiska** – liczba składowych głównych dobierana jest na podstawie wizualnej oceny wykresu osypiska dokonanej przez badacza. Na wykresie tym przedstawione są kolejne wartości własne (uszeregowana w kolejności malejącej). Wyboru dokonuje się poprzez stwierdzenie, dla której wartości składowych głównych następuje „wyprostowanie" łamanej powstałej w wyniku połączenia kolejnych punktów na wykresie. Wyprostowanie łamanej następuje w punkcie, w którym linia łącząca punkty przechodzi z „malejącej" w „stałą" – jest to miejsce końca osypywania się informacji o zmiennych oryginalnych, jaką przekazują składowe główne.



Rysunek 2.4. Wykres osypiska

Źródło: opracowanie własne.

Test istotności χ^2 – badane jest dopasowanie odtworzonej macierzy korelacji do macierzy obserwowanej.

- Czynnik przyspieszenia AF – parametr ten dla każdej wartości własnej wyznacza prędkość, z jaką krzywa opada w tym punkcie. Obliczana ona jest ze wzoru:

$$AF(i) = e(i+1) - 2e(i) + e(i-1), \qquad (2.6)$$

gdzie *e(i)* są kolejnymi wartościami własnymi. Liczba składowych głównych określona jest przez numer składowej, dla której AF osiąga największą wartość pomniejszoną o 1. Warunkiem koniecznym jest spełnienie kryterium analizy równoległej.

Składowe główne mogą zostać wyliczone na dwa sposoby – w oparciu o macierz korelacji lub macierz kowariancji. W przypadku wyboru macierzy korelacji składowe główne obliczane są jako kombinacja liniowa wystandaryzowanych zmiennych oryginalnych. Jeżeli składowe główne mają powstać w oparciu o macierz kowariancji, to wyliczane są jako kombinacja liniowa wycentrowanych względem średniej zmiennych oryginalnych.

Uzyskane tymi metodami składowe główne stanowią nowe nieskorelowane ze sobą zmienne. Oznacza to, że nie ma między nimi liniowej zależności – nie wyklucza to jednak zależności nieliniowej. Z tego względu metoda składowych głównych najlepsze efekty przynosi w przypadku zmiennych podlegających wielowymiarowemu rozkładowi normalnemu lub

rozkładów zbliżonych do normalnego. Techniką, która pozwala rozwiązać ten problem jest analiza składowych niezależnych. Rozwiązanie to dopuszcza składowe niezależne statystycznie.

Zmiana skali zmiennych pierwotnych w przypadku analizy PCA powoduje zmianę wyników (wartości składowych głównych). Z tego powodu wyniki uzyskane w oparciu o macierze korelacji i kowariancji różnią się. Jeżeli występują duże różnice w wariancjach lub cechy mierzone są na różnych skalach zaleca się wykonanie skalowania danych.

2.3.3. Analiza czynnikowa

Analiza czynnikowa jest zbiorem metod i procedur statystycznych pozwalających na badanie wzajemnych zależności pomiędzy dużą liczbą cech i wykrywanie ukrytych uwarunkowań [25]. Podstawowym celem analizy czynnikowej, jest podobnie jak w przypadku analizy składowych głównych, redukcja wymiarowości danych. Metoda ta grupuje zmienne silnie skorelowane i na ich podstawie wyznacza mniejszą liczbę czynników. Obie metody dążą do takiego samego celu. Pomimo że w praktyce wyniki otrzymywane przy pomocy wspomnianych metod są zbliżone, to nie należy ich traktować jako dwóch wariantów tej samej metody, ale jako dwie odmienne metody opierające się na różnych założeniach. Istnieje kilka znaczących różnic pomiędzy nimi – najważniejsze z nich zostały przedstawione w tabeli 2.4.

W analizie czynnikowej wartości czynników nie są określone jednoznacznie. W przypadku tej metody istnieje nieskończenie wiele rozwiązań prowadzących do identycznych powiązań pomiędzy oryginalnymi zmiennymi i czynnikami. Nowe rozwiązania uzyskuje się poprzez rotację czynników. Pozwala to również na czytelniejszą analizę otrzymanych nowych zmiennych (czynników).

Tabela 2.4. Porównanie analizy czynnikowej i analizy składowych głównych

Analiza czynnikowa	Analiza składowych głównych
Badana jest tylko część wspólna wariancji (wariancja dzielona jest wariancję wspólną (z innymi zmiennymi) oraz wariancję swoistą (charakterystyczną dla danej zmiennej)).	Badana jest całkowita wariancja.
Macierze korelacji i kowariancji są takie same (metoda ta jest niezmiennicza ze względu na skalowanie).	Macierze korelacji i kowariancji różnią się.
Dodanie kolejnych czynników do rozwiązania zmienia pozostałe.	Dodanie kolejnych składowych nie zmienia już istniejących.
Zmienna oryginalna jest funkcją czynników (wspólnych i swoistych).	Główna składowa jest funkcją zmiennych oryginalnych
Analiza pozwala na identyfikację ukrytych zmiennych (poprzez rotację struktury czynnikowej) oraz redukcję wymiarowości.	Głównym celem analizy jest redukcja wymiarowości.
Czynniki mogą być skorelowane.	Czynniki nigdy nie są ze sobą skorelowane.

Źródło: opracowanie własne.

2.3.4. Skalowanie wielowymiarowe

Skalowanie wielowymiarowe (SWW) jest, w odniesieniu do wielowymiarowych technik eksploracyjnej analizy danych, jedną z alternatyw dla analizy czynnikowej i analizy składowych głównych. Celem tego rozwiązania jest wykrycie istotnych ukrytych wymiarów pozwalających na wyjaśnienie podobieństw lub odmienności (odległości) pomiędzy badanymi procesami, obiektami, zjawiskami. Metoda SWW opiera się na macierzy niepodobieństwa między danymi wejściowymi. Zmierza ona do uporządkowania badanych obiektów poprzez wyznaczenie ich w nowym układzie współrzędnych tak, aby odległości między danymi w przestrzeni oryginalnej i transformowanej były do siebie jak najbardziej zbliżone. Z rozbieżnością w tych odległościach wiąże się jedna z najpowszechniejszych metod oceny poprawności dopasowania – stres. Wyrażony jest on wzorem:

$$Phi = \sum(d_{ij} - f(\delta_{ij}))^2, \qquad (2.7)$$

gdzie d_{ij} oznacza odtworzone odległości przy zadanej liczbie pomiarów, a δ_{ij} jest miarą określającą odległości obserwowane (dane wejściowe).

Celem algorytmu SWW jest minimalizacja tej funkcji. Końcowa wartość stresu może służyć jako miara jakości uzyskanego odwzorowania. Jedna z przyjętych metod oceny odwzorowania ze względu na funkcję stresu przedstawiona została w tabeli 2.5.

Tabela 2.5. Interpretacja współczynnika stresu

Stres	Stopień dopasowania
$\boldsymbol{stres \geq 0,2}$	Słaby
$\boldsymbol{0,1 \geq stres < 0,2}$	Przeciętny
$\boldsymbol{0,05 \geq stres < 0,1}$	Dobry
$\boldsymbol{0,025 \geq stres < 0,05}$	Doskonały
$\boldsymbol{0 \geq stres < 0,025}$	Idealny

Źródło: opracowanie własne.

Skalowanie wielowymiarowe można podzielić na dwa typy:

1. **Metryczne** – celem tego typu analizy jest minimalizacja sumy modułów (kwadratów) różnic między odległościami w otrzymanym (nowym) układzie współrzędnych, a odległościami pierwotnymi. W przypadku, gdy na wejściu analizy jest zbiór danych pierwotnych (a nie macierz niepodobieństw/odmienności) skalowanie wielowymiarowe i analiza składowych głównych (oparta o macierz kowariancji) daje takie same rezultaty. SWW określa się wtedy mianem klasycznego skalowania wielowymiarowego lub analizą współrzędnych głównych. Ten typ skalowania wykorzystywany jest w przypadku występowania w zbiorze pierwotnym tylko cech ilościowych.
2. **Niemetryczne** – celem niemetrycznego skalowania wielowymiarowego jest odwzorowanie porządku między odległościami. W tym przypadku odległość (jej wartość) nie jest brana pod uwagę. W przypadku SWW niemetrycznego dane mogą przyjmować postać ilościową i jakościową. Procedura znajdowania minimalnej wartości funkcji stresu jest iteracyjna i wymaga początkowej konfiguracji punktów.

Zaletą skalowania wielowymiarowego jest możliwość analizowania dowolnego rodzaju macierzy odległości lub podobieństwa. Macierze te mogą

reprezentować oceny podobieństwa obiektów, zjawisk, procesów, procentową zgodność między badanymi zbiorami. Metoda skalowania wielowymiarowego często wykorzystywana jest w badaniach psychologicznych, w których analizowane są ludzkie preferencje na podstawie różnego rodzaju cech, np. zbioru odpowiedzi z ankiety.

2.3.5. Analiza korespondencji

Analiza korespondencji to opisowa technika analizy tablic wielodzielnych pozwalająca na graficzne przedstawienie w niskowymiarowej przestrzeni danych w niej zawartych. Jednym z celów tej metody jest redukcja wymiarowości danych. Metoda ta pozwala na analizę struktury zmiennych jakościowych tworzących tablicę. Najczęściej używana jest do analizy dwuwymiarowych tablic kontyngencji (rozkładów dwóch łącznych cech mierzonych na skalach nominalnych) [26].

Technika analizy korespondencji dokonuje rzutowania oryginalnych danych na przestrzeń o jak najmniejszym wymiarze. Istotnym czynnikiem charakteryzującym dokładność odwzorowania jest inercja. Odzwierciedla ona, jaką część wariancji objaśnia zmienna z nowego układu współrzędnych. Przed rozpoczęciem procesu analizy należy zbadać, czy występuje zależność pomiędzy zmiennymi występującymi w tablicy kontyngencji, np. wykorzystując test niezależności chi-kwadrat.

2.4. Adaptacja danych medycznych

Wykorzystując dwie dostępne bazy danych autor zdecydował się na badania nad odpowiednim przygotowaniem danych wejściowych. Jest to istotne z punktu widzenia projektu, w którym przeprowadzana będzie analiza wielu danych w trybie rzeczywistym. W dostępnej autorowi literaturze analiza składowych głównych w aplikacjach medycznych stosowana jest wyłącznie do analizy sygnału EKG.

Autor w niniejszej pracy proponuje zastosowanie analizy korespondencji i analizy składowych głównych w przypadku wielorakiego rodzaju parametrów fizjologicznych. Są to dane wyrażone zarówno na skali proporcjonalnej, jak i nominalnej.

W przypadku danych jakościowych autor przeprowadził ich dekompozycję. Pomimo że sieci neuronowe są w stanie rozwiązywać złożone problemy nieliniowe to zastosowanie binaryzacji sygnałów jakościowych poprawiło skuteczność działania sieci. Może to być związane z błędnym (przypadkowym) kodowaniem zmiennych jakościowych przez lekarzy. W takim przypadku mogą pojawiać się znaczne nieliniowości, które mogą utrudniać poprawne rozpoznanie przez sieć neuronową.

Binaryzacja danych jest często wykonywanym procesem, jeżeli w zbiorze danych uczących sieci neuronowej występują zmienne na skali nominalnej. Jak można zauważyć z badań przeprowadzonych przez autora, dekompozycja danych podnosi skuteczność predykcyjną sieci.

W przypadku danych z bazy w Cleveland, w której występowało stosunkowo dużo danych jakościowych przeprowadzono porównanie wyników skuteczności działania sieci w przypadku danych oryginalnych oraz danych po dekompozycji. Operacja ta spowodowała zwiększenie się liczby danych wejściowych z 13 do 25. Dla lepszej generalizacji wniosków każdy rodzaj sieci przebadany został 100 razy (wykorzystując różne wagi początkowe). Przebadane zostały sieci o następujących metodach uczenia:

- największego spadku;
- gradientów sprzężonych;
- algorytmu zmiennej metryki (BFGS).

Każda z metod uczenia sieci przebadana została z dwiema różnymi funkcjami aktywacji – logistycznej oraz tangensa hiperbolicznego.

Wszystkie przedstawione badania zostały przeprowadzone dla losowo podzielonego zbioru danych. Przebadane zostały przypadki, w których zbiór uczący wynosił 50%, 60%, 70% oraz 80%. Wszystkie porównywane ze sobą skuteczności działania sieci odnoszą się do dokładnie takiego samego zbioru uczącego. W uczeniu sieci brał udział również zbiór testowy. Służył on weryfikacji algorytmu uczenia i pozwalał na zatrzymanie procesu uczenia w przypadku zwiększania się liczby błędnie rozpoznanych wzorców w zbiorze testowym. Był on zabezpieczeniem procesu uczenia przed nadmiernym dopasowaniem się zbioru uczącego (przeuczeniem).

Porównanie skuteczności sieci przed binaryzacją i po binaryzacji zostało zilustrowane w tabelach 2.6-2.8.

Tabela 2.6. Skuteczność sieci przed i po binaryzacji uczonej metodą największego spadku – średnia dla 100 przypadków uczenia

	Zbiór uczący 50%		Zbiór uczący 60%		Zbiór uczący 70%		Zbiór uczący 80%	
Liczba neuronów w warstwie ukrytej	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji
4	81,53	82,43	83,15	83,05	78,50	83,38	68,96	82,76
6	81,53	82,43	82,89	83,05	78,72	83,68	68,96	82,76
8	81,53	82,43	82,44	83,05	78,72	84,01	68,96	82,76
10	81,53	82,43	82,64	83,05	78,83	84,05	68,96	82,76
12	81,53	82,43	83,24	83,05	78,95	84,23	68,96	82,76
14	81,53	82,43	83,45	83,05	79,22	84,21	68,96	82,76
16	81,53	82,43	83,35	83,05	79,41	84,31	68,96	82,76
18	81,53	82,43	82,79	83,05	79,11	84,15	68,96	82,76
20	81,53	82,43	82,06	83,05	78,81	83,96	68,96	82,76
22	81,53	82,43	82,18	83,05	78,86	84,03	68,96	82,76
24	81,53	82,43	82,64	83,05	78,86	83,94	68,96	82,76

Źródło: opracowanie własne.

W tabeli 2.6 zebrano skuteczność uczenia sieci neuronowej metodą największego spadku dla czterech rozmiarów zbioru uczącego (50%, 60%, 70%, 80%) oraz dla jedenastu rozmiarów warstwy ukrytej (od 4 do 24 neuronów w warstwie ukrytej). Dla 95,5% wszystkich przypadków przedstawionych w tabeli można zauważyć poprawę skuteczności sieci po dekompozycji parametrów wejściowych. Trzy wyjątki od tej reguły dotyczą sieci uczonej zbiorem 60% i mającej 12, 14 i 16 neuronów w warstwie ukrytej. Można zaobserwować prawidłowość, że przyrost skuteczności jest tym większy im większy jest zbiór uczący. Najmniejszą poprawę odnotowano dla przypadków zbioru uczącego 50% oraz 60%. – poprawa o około 1%. Dla przypadku zbioru uczącego 70% zaobserwowano poprawę w granicach 5%. Dla największego zbioru uczącego skuteczność uczenia

zwiększyła się o blisko 14%. Zgodnie z literaturą metoda ta jest najmniej odporna na zatrzymywanie się w lokalnych minimach spośród wszystkich metod uczenia prezentowanych w niniejszej pracy. Znalazło to potwierdzenie w przeprowadzonych badaniach – dla zbioru uczącego 60% (sieć po binaryzacji) oraz 80% (przed i po binaryzacji) proces uczenia zatrzymywał się w lokalnym minimum. Sytuacja taka miała miejsce dla wszystkich (100) przypadków uczenia, niezależnie od liczby neuronów w warstwie ukrytej.

Tabela 2.7. Skuteczność sieci przed i po binaryzacji uczonej metodą gradientów sprzężonych – średnia dla 100 przypadków uczenia

	Zbiór uczący 50%		Zbiór uczący 60%		Zbiór uczący 70%		Zbiór uczący 80%	
Liczba neuronów w warstwie ukrytej	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji
4	79,76	83,98	80,68	84,33	77,82	82,86	69,51	80,06
6	79,85	84,11	80,78	84,89	77,71	83,10	68,95	81,77
8	79,76	83,98	80,81	84,57	77,60	83,27	69,32	82,41
10	79,70	84,54	80,85	85,78	77,52	83,43	69,11	81,89
12	80,19	84,25	80,78	86,27	77,36	82,91	69,24	81,33
14	79,76	84,88	80,74	85,65	77,41	82,34	69,10	81,75
16	79,41	84,57	81,05	85,98	77,41	81,41	68,75	82,62
18	79,23	84,73	81,08	85,83	77,21	81,19	69,58	82,21
20	79,08	84,65	81,22	84,88	77,09	81,13	69,51	81,65
22	79,03	84,02	80,99	84,93	76,77	80,86	69,44	81,59
24	79,22	83,38	80,91	84,72	76,46	80,95	69,37	81,86

Źródło: opracowanie własne.

Tabela 2.7 prezentuje skuteczność sieci uczonej metodą gradientów sprzężonych. Wartości zebrane w tabeli są średnimi dla 100 przypadków uczenia. W badaniach przedstawionych w tabeli można zauważyć, że dla każdego przypadku binaryzacja powoduje wzrost skuteczności uczenia.

Najbardziej wyraźna poprawa występuje dla zbioru uczącego 80% – wzrost skuteczności ok. 11-12%. Dla pozostałych zbiorów uczących wzrost skuteczności osiągnął ok. 4-5%. Zbiór uczący 60% pozwolił na osiągnięcie największej bezwzględnej skuteczności uczenia – dotyczy to zarówno sieci przed, jak i po binaryzacji. Dla sieci przed binaryzacją widać zdecydowany spadek skuteczności dla zbiorów uczących powyżej 70%.

Tabela 2.8. Skuteczność sieci uczonej przed i po binaryzacji metodą algorytmu BFGS – średnia dla 100 przypadków uczenia

	Zbiór uczący 50%		Zbiór uczący 60%		Zbiór uczący 70%		Zbiór uczący 80%	
Liczba neuronów w warstwie ukrytej	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji	Skuteczność sieci przed binaryzacją	Skuteczność sieci po binaryzacji
4	80,54	83,16	80,68	84,33	77,82	82,86	69,51	80,06
6	80,67	82,88	80,78	84,89	77,71	83,10	68,95	81,77
8	80,84	82,54	80,81	84,57	77,60	83,27	69,32	82,41
10	80,56	83,16	80,85	85,78	77,52	83,43	69,11	81,89
12	80,33	83,67	80,78	86,27	77,36	82,91	69,24	81,33
14	80,45	83,75	80,74	85,65	77,41	82,34	69,10	81,75
16	80,45	83,84	81,05	85,98	77,41	81,41	68,75	82,62
18	80,33	84,32	81,08	85,83	77,21	81,19	69,58	82,21
20	80,06	84,27	81,22	84,88	77,09	81,13	69,51	81,65
22	80,15	83,94	80,99	84,93	76,77	80,86	69,44	81,59
24	80,00	84,05	80,91	84,72	76,46	80,95	69,37	81,86

Źródło: opracowanie własne.

Tabela 2.8 zawiera zestawienie skuteczności sieci uczonej za pomocą algorytmu zmiennej metryki. Podobnie jak w przypadku uczenia metodą gradientów sprzężonych dla 100% przypadków binaryzacja poprawiła skuteczność działania sieci. Optymalny zbiór uczący dla omawianego zestawu danych w przypadku algorytmu BFGS wynosi 60%. Zarówno dla

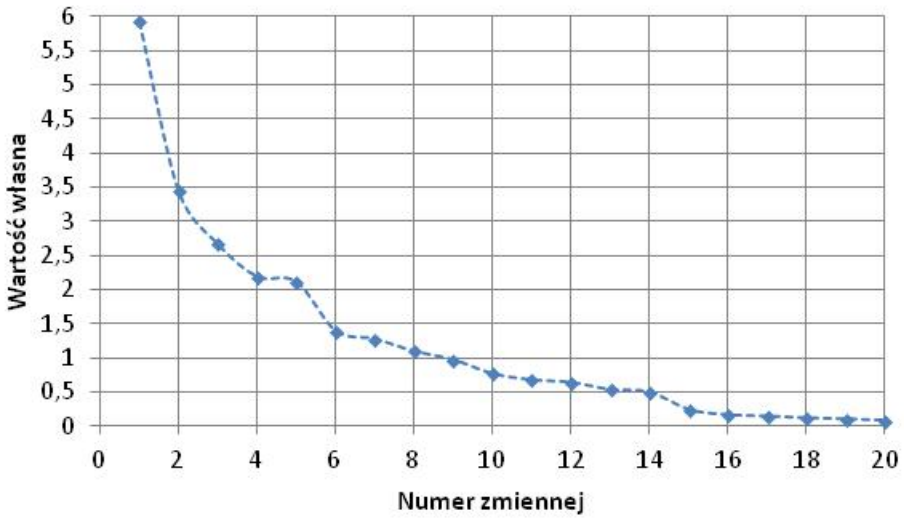
sieci przed binaryzacją, jak i po binaryzacji dla takiego zbioru uczącego osiągnięta największą skuteczność sieci neuronowej – odpowiednio ok. 81% (przed binaryzacją), ok. 86% (po binaryzacji). Największy przyrost skuteczności można zauważyć dla zbioru uczącego 80% (ok. 12%). Pozostałe przypadki charakteryzują się wzrostem na poziomie ok. 4-5%. Jest to sytuacja zbliżona do zaobserwowanej podczas uczenia sieci metodą gradientów sprzężonych

Jak widać w zamieszczonych tabelach binaryzacja zmiennych wyrażonych na skalach nominalnych pozytywnie wpływa (zwiększa skuteczności) na algorytm klasyfikacji. Dotyczy to wszystkich metod uczenia. Najniższa bezwzględna skuteczność sieci występuje w przypadku metody największego spadku. Może to być wytłumaczone przez wolniejszą zbieżność tej metody oraz tendencje do utykania w lokalnym minimum. Pozostałe metody lepiej radzą sobie z nieliniowościami i posiadają lepsze mechanizmy szukania globalnego minimum.

2.5. Redukcja parametrów medycznych

W celu ograniczenia liczby parametrów wejściowych sieci neuronowej została zastosowana metoda konstrukcji nowych cech w oparciu o ideę składowych głównych. W dostępnej autorowi literaturze istnieje kilka pozycji opisujących zastosowanie tej metody do redukcji danych charakteryzujących przebieg EKG [27, 28, 29]. Autor proponuje wykorzystanie tej metody do redukcji różnego rodzaju parametrów pochodzących z typowego monitora funkcji życiowych. Oprócz parametrów z przebiegu EKG występują tam również informacje na temat składu chemicznego krwi oraz innych parametrów charakteryzujących pracę serca i ogólny stan pacjenta.

W celu redukcji danych wejściowych modułu decyzyjnego (sieci neuronowej) został skonstruowany zredukowany zbiór zawierający liniowo niezależne parametry wejściowe. Na rysunku 2.5 przedstawiony został wykres osypiska prezentujący odsetek całkowitej wariancji wyjaśnianej przez odpowiednią liczbę nowych parametrów.



Rysunek 2.5. Wykres osypiska

Źródło: opracowanie własne.

Tabela 2.9 prezentuje wartości własne i skumulowaną wartość wariancji dla odpowiedniej liczby parametrów.

Tabela 2.9. Skuteczność odwzorowania zmiennych pierwotnych dla bazy danych Cleveland

Numer	Wartość własna nowej cechy	Wariancja poszczególnych nowych cech (%)	Skumulowana jakość odwzorowania (%)
1	5,93	23,73	23,73
2	3,44	13,76	37,49
3	2,67	10,66	48,15
4	2,19	8,75	56,91
5	2,11	8,45	65,37
6	1,39	5,57	70,94
7	1,27	5,09	76,04
8	1,09	4,37	80,42
9	0,97	3,87	84,29

Numer	Wartość własna nowej cechy	Wariancja poszczególnych nowych cech (%)	Skumulowana jakość odwzorowania (%)
10	0,76	3,05	87,34
11	0,67	2,69	90,04
12	0,64	2,54	92,58
13	0,53	2,12	94,71
14	0,49	1,97	96,68
15	0,24	0,95	97,64
16	0,16	0,63	98,27
17	0,14	0,56	98,83
18	0,11	0,45	99,29
19	0,10	0,4	99,69
20	0,08	0,31	100,0

Źródło: opracowanie własne.

Jak widać zarówno z tabeli, jak i wykresu osypiska, zastosowanie 20 z 25 nowych liniowo niezależnych składowych gwarantuje zachowanie jakości odwzorowania na poziomie 100%. Należy również zauważyć, że w przypadku ograniczenia liczby nowych parametrów do wartości sprzed dekompozycji zachowana skumulowana wariancja przekracza 90%. Istnieje zatem duże prawdopodobieństwo poprawnego odwzorowania struktury procesu przy pomocy jedynie połowy zmiennych. Badania sprawdzające skuteczność działania sieci neuronowej dla różnej liczby nowych niezależnych liniowo parametrów zostały przeprowadzone i umieszczone w tabelach 2.12-2.14.

Z tabeli 2.9 jednoznacznie widać, że za pomocą jedynie 20 z 25 zmiennych możemy wyrazić całkowitą wariancję (czyli 100% jakości odwzorowania). Oznacza to, że nie wszystkie pierwotne cechy niosły ze sobą informację użyteczną. Dzięki temu można zredukować 5 cech pierwotnych, które były kombinacją liniową pozostałych.

W tabeli 2.10 zaprezentowana została korelacja pomiędzy pierwotnymi czynnikami a nowymi. Można z niej odczytać, jak intensywna jest korelacja poszczególnych czynników ze zmiennymi pierwotnymi.

Tabela 2.10. Tabela korelacji między nowymi i pierwotnymi cechami

	Czynn. 1	Czynn. 2	Czynn. 3	Czynn. 4	Czynn. 5	Czynn. 6	Czynn. 7	Czynn. 8	Czynn. 9	Czynn. 10
Age	-0,288	0,121	0,013	-0,203	0,218	-0,08	-0,23	0,048	-0,062	0,067
Sex	-0,311	-0,195	-0,396	-0,096	-0,406	0,345	0,36	0,466	0,197	-0,117
Cp1	-0,649	-0,068	-0,144	0,507	0,375	-0,213	0,043	0,042	0,269	0,112
Cp2	0,383	0,035	0,236	-0,635	-0,294	-0,493	0,124	0,046	0,04	-0,124
Cp3	0,387	-0,012	-0,109	0,102	-0,013	0,719	-0,125	-0,264	-0,348	-0,052
Cp4	0,031	0,085	0,023	-0,022	-0,189	0,225	-0,115	0,213	-0,086	0,073
Resting bloo	-0,147	0,115	-0,023	0,007	-0,075	-0,095	-0,21	0,049	-0,19	0,14
Chloesterol	-0,076	0,162	-0,007	0,003	0,089	-0,115	-0,129	-0,179	-0,077	-0,032
Sugar	-0,037	0,069	-0,098	-0,222	-0,083	-0,138	-0,293	0,439	-0,421	0,326
ECG1	-0,306	0,907	-0,198	0,011	-0,163	-0,002	0,006	-0,051	0,002	-0,039
ECG2	-0,057	0,017	0,113	0,024	0,073	-0,058	0	-0,025	-0,009	0,126
ECG3	0,319	-0,911	0,172	-0,016	0,146	0,016	-0,006	0,057	0	0,01
Heart Rate	0,532	0,035	-0,191	0,031	-0,18	0,047	-0,102	-0,053	0,035	-0,101
Angina	-0,582	-0,101	-0,038	0,207	0,184	-0,213	0,43	0,049	-0,53	-0,223
Oldpeak	-0,513	-0,019	0,167	0,002	-0,062	-0,092	-0,114	0,171	-0,083	0,162
SlopeST1	-0,071	0,026	-0,03	0,094	-0,17	-0,103	-0,03	0,226	-0,254	0,604
SlopeST2	-0,608	0,055	0,717	-0,101	0,008	0,185	0,005	-0,032	0,09	-0,163
SlopeST3	0,643	-0,068	-0,7	0,053	0,079	-0,131	0,01	-0,084	0,041	-0,148
Vessel1	-0,21	0,051	-0,028	-0,097	0,014	-0,033	-0,183	0,131	0,126	0,077
Vessel2	-0,25	-0,024	-0,136	0,014	0,268	-0,047	-0,63	0,332	-0,08	-0,446
Vessel3	-0,258	-0,004	-0,212	-0,614	0,418	0,163	0,388	-0,249	0,015	0,235
Vessel4	0,492	-0,006	0,284	0,556	-0,539	-0,089	0,191	-0,08	-0,023	0,062
Thal1	-0,682	-0,365	-0,259	-0,082	-0,4	-0,102	-0,156	-0,319	-0,021	-0,033
Thal2	-0,173	0,022	0,068	0,017	0,001	0,157	0,001	0,382	-0,012	0,202
Thal3	0,751	0,347	0,221	0,072	0,392	0,025	0,152	0,129	0,027	-0,065

Źródło: opracowanie własne.

Redukcja danych wejściowych zastosowana została również dla drugiej bazy danych. Z uwagi na fakt, że wszystkie parametry znajdujące się w bazie danych Łódź były ciągłe, nie zastosowano przy przygotowaniu

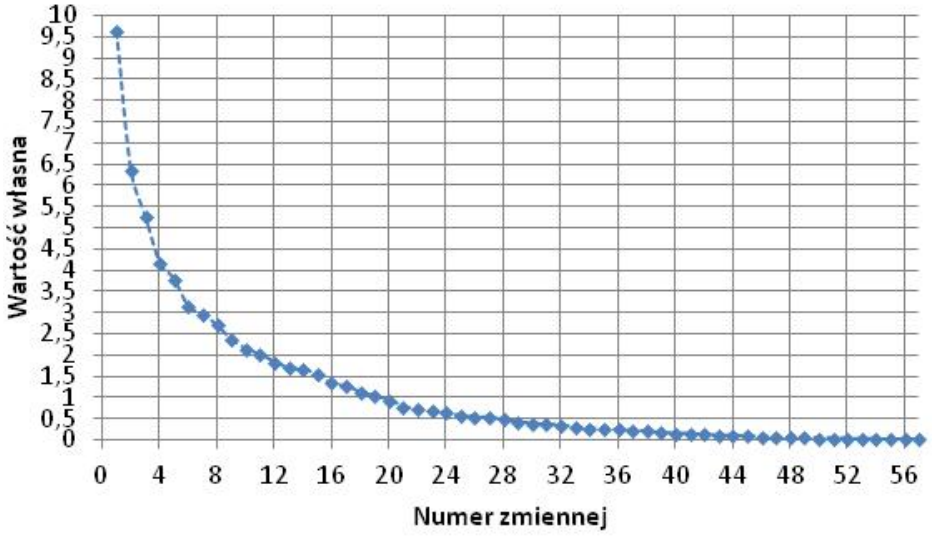
danych procesu dekompozycji. W przypadku bazy danych z Łodzi zaimplementowano jedynie metodę redukcji danych liniowo zależnych. W ten sposób uzyskano 57 nowych parametrów odpowiedzialnych za 100% użytecznej informacji. Wyniki te zostały przedstawione w tabeli 2.11 oraz na rysunku 2.6.

Tabela 2.11. Skuteczność odwzorowania zmiennych pierwotnych dla bazy danych Łódź

Numer	Wartość własna nowej cechy	Wariancja poszczególnych nowych cech (%)	Skumulowana jakość odwzorowania (%)
1	9,65	14,4	14,4
2	6,38	9,52	23,93
3	5,29	7,89	31,82
4	4,19	6,26	38,08
5	3,80	5,67	43,76
6	3,16	4,72	48,42
7	2,96	4,42	52,96
8	2,71	4,05	56,96
9	2,38	3,55	60,51
10	2,16	3,22	63,74
11	2,02	3,01	66,75
12	1,85	2,76	69,52
13	1,70	2,54	72,06
14	1,66	2,48	74,55
15	1,55	2,32	76,88
16	1,35	2,01	78,89
17	1,27	1,89	80,78
18	1,14	1,7	82,49
19	1,04	1,55	84,05
20	0,95	1,42	85,47
21	0,80	1,19	86,66
22	0,73	1,09	87,75
23	0,69	1,03	88,76
24	0,65	0,97	89,73
25	0,58	0,86	90,6
26	0,56	0,83	91,43
27	0,53	0,79	92,23
28	0,49	0,73	92,97
29	0,44	0,65	93,62
30	0,40	0,59	94,21
31	0,38	0,56	94,77
32	0,36	0,53	95,29
33	0,32	0,48	95,77

34	0,28	0,42	96,2
35	0,27	0,41	96,62
36	0,25	0,38	97
37	0,23	0,35	97,37
38	0,21	0,32	97,69
39	0,18	0,27	97,97
40	0,16	0,24	98,22
41	0,16	0,233	98,47
42	0,14	0,204	98,68
43	0,12	0,178	98,87
44	0,11	0,171	99,05
45	0,10	0,146	99,2
46	0,08	0,126	99,33
47	0,08	0,112	99,45
48	0,06	0,095	99,55
49	0,05	0,082	99,63
50	0,05	0,076	99,71
51	0,05	0,074	99,77
52	0,04	0,063	99,82
53	0,03	0,047	99,86
54	0,03	0,045	99,91
55	0,02	0,037	99,94
56	0,02	0,031	99,97
57	0,02	0,026	100

Źródło: opracowanie własne.



Rysunek 2.6. Wykres osypiska dla bazy danych z Łodzi

Źródło: opracowanie własne.

2.6. Sieć neuronowa po redukcji

Po przeprowadzeniu redukcji danych i wybraniu 13 parametrów z uzyskanych po dekompozycji 25 przeprowadzone zostały badania skuteczności działania sieci neuronowej. Należy zauważyć, że dla 20 z 25 parametrów jakość odwzorowania struktury danych wynosi 100%. W tabelach 2.12-2.14 porównane zostały wyniki działania sieci neuronowej po wykonaniu redukcji danych wejściowych ze skutecznością sieci dla zmiennych wejściowych zbinaryzowanych.

Tabela 2.12. Porównanie skuteczności uczenia sieci po binaryzacji oraz po binaryzacji i redukcji metodą największego spadku – 100 przypadków uczenia

	Zbiór uczący 50%		Zbiór uczący 60%		Zbiór uczący 70%		Zbiór uczący 80%	
Liczba neuronów w warstwie ukrytej	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji
4	82,43	82,32	83,05	84,08	83,38	85,50	82,76	92,31
6	82,43	82,32	83,05	84,52	83,68	85,40	82,76	92,31
8	82,43	82,32	83,05	85,03	84,01	85,43	82,76	92,31
10	82,43	82,32	83,05	84,95	84,05	84,79	82,76	92,31
12	82,43	82,32	83,05	85,16	84,23	85,15	82,76	92,31
14	82,43	82,32	83,05	84,88	84,21	85,66	82,76	92,31
16	82,43	82,32	83,05	84,67	84,31	86,50	82,76	92,31
18	82,43	82,32	83,05	84,57	84,15	86,32	82,76	92,31
20	82,43	82,32	83,05	84,53	83,96	85,95	82,76	92,31
22	82,43	82,32	83,05	84,61	84,03	85,42	82,76	92,31
24	82,43	82,32	83,05	84,67	83,94	85,10	82,76	92,31

Źródło: opracowanie własne.

W tabeli 2.12 zestawione zostały skuteczności uczenia dla czterech wielkości zbioru uczącego metodą największego spadku. Dla trzech przypadków redukcja liniowych zależności przyniosła poprawę skuteczności działania sieci. Jedynie w przypadku zbioru uczącego 50% sieć po redukcji osiągnęła nieznacznie gorszą skuteczność działania. Należy jednak zauważyć, że dla zbioru uczącego 50% proces uczenia sieci metodą największego spadku utykał w lokalnym minimum (dla wszystkich 100 przypadków uczenia). Miało to miejsce dla sieci zarówno przed redukcją, jak i po redukcji, niezależnie od liczby neuronów w warstwie ukrytej. Największy wzrost skuteczności sieci zaobserwowano dla zbioru uczącego 80%. W tym przypadku proces nauki również zatrzymywał się w lokalnym minimum (przed redukcją oraz po redukcji).

Tabela 2.13. Porównanie skuteczności uczenia sieci po binaryzacji oraz po binaryzacji i redukcji metodą gradientów sprzężonych – 100 przypadków uczenia

	Zbiór uczący 50%		Zbiór uczący 60%		Zbiór uczący 70%		Zbiór uczący 80%	
Liczba neuronów w warstwie ukrytej	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji
4	83,98	92,51	84,33	90,87	82,86	89,75	80,06	92,76
6	84,11	92,77	84,89	91,11	83,10	90,30	81,77	92,67
8	83,98	92,83	84,57	91,32	83,27	90,95	82,41	92,54
10	84,54	93,34	85,78	91,24	83,43	91,35	81,89	92,46
12	84,25	93,01	86,27	91,05	82,91	90,85	81,33	92,39
14	84,88	92,54	85,65	91,44	82,34	90,20	81,75	92,46
16	84,57	93,85	85,98	91,77	81,41	89,80	82,62	92,46
18	84,73	93,76	85,83	91,87	81,19	89,65	82,21	92,43
20	84,65	94,09	84,88	92,30	81,13	89,85	81,65	92,39
22	84,02	93,85	84,93	91,55	80,86	89,50	81,59	92,33
24	83,38	93,64	84,72	91,06	80,95	89,35	81,86	92,31

Źródło: opracowanie własne.

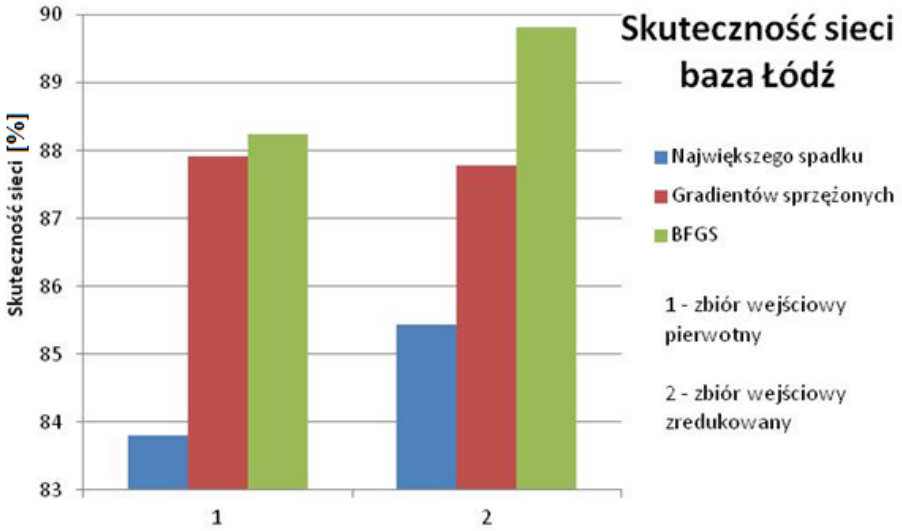
Tabela 2.13 przedstawia porównanie skuteczności uczenia sieci przed dekompozycją i po dekompozycji dla procesu uczenia metodą gradientów sprzężonych. Dla wszystkich zbiorów uczących redukcja liniowych zależności przyniosła poprawę skuteczności uczenia. Największy przyrost skuteczności wystąpił dla zbioru uczącego 80%. W przypadku sieci po redukcji i binaryzacji, gdy zbiorem uczącym była połowa całego zbioru danych, uzyskano największą skuteczność uczenia. W porównaniu do procesu uczenia metodą największego spadku, dla wszystkich przypadków, oprócz jednego, zaobserwowano poprawę skuteczności działania sieci. Jedynym wyjątkiem jest przypadek zbioru uczącego 80% oraz 24 neuronów w warstwie ukrytej, gdzie skuteczność sieci była jednakowa (92,31%).

Tabela 2.14. Porównanie skuteczności uczenia sieci po binaryzacji oraz po binaryzacji i redukcji metodą algorytmu BFGS – 100 przypadków uczenia

	Zbiór uczący 50%		Zbiór uczący 60%		Zbiór uczący 70%		Zbiór uczący 80%	
Liczba neuronów w warstwie ukrytej	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci po binaryzacji	Skuteczność sieci po binaryzacji i redukcji
4	83,16	89,64	84,33	89,85	82,86	88,25	80,06	92,23
6	82,88	89,58	84,89	89,24	83,10	87,80	81,77	92,67
8	82,54	89,73	84,57	88,83	83,27	88,25	82,41	93,03
10	83,16	89,64	85,78	88,83	83,43	89,25	81,89	92,83
12	83,67	90,07	86,27	89,24	82,91	88,75	81,33	92,69
14	83,75	90,15	85,65	89,93	82,34	88,75	81,75	92,64
16	83,84	90,78	85,98	90,94	81,41	88,25	82,62	92,53
18	84,32	90,46	85,83	90,55	81,19	87,00	82,21	92,64
20	84,27	90,33	84,88	90,27	81,13	87,25	81,65	92,69
22	83,94	90,55	84,93	90,55	80,86	87,75	81,59	92,69
24	84,05	90,91	84,72	90,33	80,95	88,25	81,86	92,77

Źródło: opracowanie własne.

W tabeli 2.14 zebrano wyniki skuteczności działania sieci po dekompozycji oraz po dekompozycji i redukcji dla procesu uczenia metodą algorytmu zmiennej metryki (BFGS). Dla wszystkich przypadków (uwzględniając liczbę neuronów w warstwie ukrytej oraz rozmiar zbioru uczącego) zaobserwowano poprawę skuteczności działania sieci, podobnie jak w przypadku uczenia metodą gradientów sprzężonych. Największa poprawa oraz największa bezwzględna wartość skuteczności uczenia uzyskana została dla zbioru uczącego 80%.



Rysunek 2.7. Porównanie skuteczności uczenia sieci przed redukcją i po redukcji dla bazy danych z Łodzi

Źródło: opracowanie własne.

Podobne obliczenia przeprowadzone zostały również dla bazy danych z Łodzi. Na rysunku 2.7 przedstawione zostały porównania skuteczności działania sieci dla trzech stosowanych algorytmów uczenia (wyniki uśrednione dla 100 przypadków uczenia). Do zbioru zredukowanego użyto 32 nowych zmiennych – odpowiada to 95.29% użytecznej informacji. Wartość ta została wybrana po rozmowach z lekarzami ze Szpitala im. Barlickiego w Łodzi, którzy na podstawie doświadczenia i literatury uznali, że minimalna wartość użytecznej informacji branej pod uwagę przy uczeniu powinna wynosić 95%. Badania zostały przeprowadzone dla przypadku, gdy liczba neuronów w warstwie ukrytej wynosiła 12. Dla tak skonstruowanej sieci otrzymywane były najwyższe skuteczności działania sieci. Można zatem uznać, że prezentowana struktura sieci neuronowej została dobrana opty-

malnie pod względem skuteczności uczenia. Dla innej (większej lub mniejszej) liczby neuronów w warstwie ukrytej otrzymano niższą jakość nauczania.

Z przedstawionej tabeli 2.9 wynika, że po zastosowaniu redukcji danych liniowo zależnych ograniczono liczbę parametrów z 25 do 20 przy zachowaniu 100% użytecznej informacji. Tabele 2.12-2.14 pokazują, że redukcja danych do 13 parametrów pozwala na uzyskanie co najmniej takiej samej jakości predykcyjnej jak przed redukcją. Podobny wniosek można wyciągnąć dla bazy danych z Łodzi. 67 parametrów pierwotnych zostało zredukowanych do 57 nowych zmiennych przy zachowaniu 100% informacji. Do nauki sieci neuronowej wykorzystano 32 nowe zmienne (co odpowiada 95,29% informacji) i, tak jak pokazuje rysunek 2.7, dla wszystkich algorytmów uczenia uzyskano skuteczność sieci przynajmniej na podobnym poziomie.

Tabela 2.15. Porównanie skuteczności sieci o 13 wejściach (pierwotnej i otrzymanej po binaryzacji i redukcji) metodą największego spadku

	Zbiór uczący 50%		Zbiór uczący 60%		Zbiór uczący 70%		Zbiór uczący 80%	
Liczba neuronów w warstwie ukrytej	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji
4	81,53	82,32	83,15	84,08	78,50	85,50	68,96	92,31
6	81,53	82,32	82,89	84,52	78,72	85,40	68,96	92,31
8	81,53	82,32	82,44	85,03	78,72	85,43	68,96	92,31
10	81,53	82,32	82,64	84,95	78,83	84,79	68,96	92,31
12	81,53	82,32	83,24	85,16	78,95	85,15	68,96	92,31
14	81,53	82,32	83,45	84,88	79,22	85,66	68,96	92,31
16	81,53	82,32	83,35	84,67	79,41	86,50	68,96	92,31
18	81,53	82,32	82,79	84,57	79,11	86,32	68,96	92,31
20	81,53	82,32	82,06	84,53	78,81	85,95	68,96	92,31
22	81,53	82,32	82,18	84,61	78,86	85,42	68,96	92,31
24	81,53	82,32	82,64	84,67	78,86	85,10	68,96	92,31

Źródło: opracowanie własne.

W tabeli 2.15 została porównana skuteczność uczenia sieci neuronowej metodą największego spadku dla danych pierwotnych i danych po dekompozycji i redukcji. Dla wszystkich rozmiarów zbioru uczącego nastąpiła poprawa skuteczności działania sieci. Z tabeli można odczytać, że zysk z przeprowadzenia dekompozycji i redukcji rośnie wraz ze zwiększaniem zbioru uczącego. Dla sieci pierwotnej największa skuteczność uczenia osiągnięta została dla zbioru uczącego 60%. W przypadku binaryzacji i redukcji największą skuteczność zaobserwowano dla zbioru uczącego 80%.

Tabela 2.16. Porównanie skuteczności sieci o 13 wejściach (pierwotnej i otrzymanej po binaryzacji i redukcji) metodą gradientów sprzężonych

	Zbiór uczący 50%		Zbiór uczący 60%		Zbiór uczący 70%		Zbiór uczący 80%	
Liczba neuronów w warstwie ukrytej	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji
4	79,76	92,51	80,68	90,87	77,82	89,75	69,51	92,76
6	79,85	92,77	80,78	91,11	77,71	90,30	68,95	92,67
8	79,76	92,83	80,81	91,32	77,60	90,95	69,32	92,54
10	79,70	93,34	80,85	91,24	77,52	91,35	69,11	92,46
12	80,19	93,01	80,78	91,05	77,36	90,85	69,24	92,39
14	79,76	92,54	80,74	91,44	77,41	90,20	69,10	92,46
16	79,41	93,85	81,05	91,77	77,41	89,80	68,75	92,46
18	79,23	93,76	81,08	91,87	77,21	89,65	69,58	92,43
20	79,08	94,09	81,22	92,30	77,09	89,85	69,51	92,39
22	79,03	93,85	80,99	91,55	76,77	89,50	69,44	92,33
24	79,22	93,64	80,91	91,06	76,46	89,35	69,37	92,31

Źródło: opracowanie własne.

Tabela 2.16 wyniki badania skuteczności działania sieci dla metody gradientów sprzężonych. Wszystkie przebadane sieci wykazywały wzrost skuteczności działania w wyniku dekompozycji i redukcji liniowych zależności. Największa poprawa skuteczności została zaobserwowana dla zbioru uczącego 80%. Najwyższa bezwzględna wartość skuteczności sieci po redukcji danych liniowo zależnych została osiągnięta dla zbioru uczącego

50%. W przypadku sieci pierwotnej sytuacja ta miała miejsce dla zbioru uczącego 60%.

Tabela 2.17. Porównanie skuteczności sieci o 13 wejściach (pierwotnej i otrzymanej po binaryzacji i redukcji) metodą algorytmu BFGS

	Zbiór uczący 50%		Zbiór uczący 60%		Zbiór uczący 70%		Zbiór uczący 80%	
Liczba neuronów w warstwie ukrytej	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji	Skuteczność sieci pierwotnej	Skuteczność sieci po binaryzacji i redukcji
4	80,54	89,64	80,68	89,85	77,82	88,25	69,51	92,23
6	80,67	89,58	80,78	89,24	77,71	87,80	68,95	92,67
8	80,84	89,73	80,81	88,83	77,60	88,25	69,32	93,03
10	80,56	89,64	80,85	88,83	77,52	89,25	69,11	92,83
12	80,33	90,07	80,78	89,24	77,36	88,75	69,24	92,69
14	80,45	90,15	80,74	89,93	77,41	88,75	69,10	92,64
16	80,45	90,78	81,05	90,94	77,41	88,25	68,75	92,53
18	80,33	90,46	81,08	90,55	77,21	87,00	69,58	92,64
20	80,06	90,33	81,22	90,27	77,09	87,25	69,51	92,69
22	80,15	90,55	80,99	90,55	76,77	87,75	69,44	92,69
24	80,00	90,91	80,91	90,33	76,46	88,25	69,37	92,77

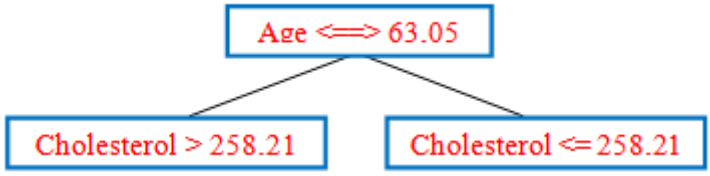
Źródło: opracowanie własne.

W tabeli 2.17 zaprezentowano porównanie wyników skuteczności procesu uczenia sieci za pomocą algorytmu zmiennej metryki (BFGS). Wykonanie dekompozycji danych, zastosowanie metody redukcji danych liniowo zależnych i odrzucenie czynników niosących najmniej informacji (tak, aby uzyskać taką samą liczbę zmiennych wejściowych) dla wszystkich przebadanych przypadków poprawiło skuteczność uczenia. Algorytm BFGS największą skuteczność (ponad 92%) uzyskał dla zbioru uczącego 80%. Dla innych rozmiarów zbioru uczącego skuteczność wynosiła ok. 90%.

Z tabeli 2.15-2.17 można wnioskować, że dekompozycja parametrów jakościowych stanu kardiologicznego pacjenta wraz z usunięciem liniowych zależności wszystkich parametrów umożliwia wzrost skuteczności rozpoznania stanu kardiologicznego przy jednoczesnym zachowaniu takiej samej liczby parametrów wejściowych.

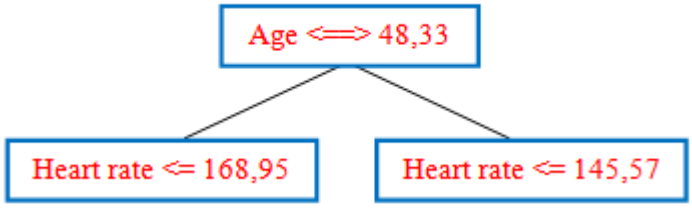
2.7. Drzewo decyzyjne

W ramach badań nad zastosowaniem metod sztucznej inteligencji do analizy ryzyka chorób kardiologicznych zostały zaprojektowane drzewa decyzyjne. Do klasyfikacji wykorzystane zostały tylko sygnały ciągłe. Na rysunkach 2.8-2.10 zaprezentowane zostały zaprojektowane proste drzewa decyzyjne. Przedstawione drzewa zostały zasymulowane w specjalnie do tego zaprojektowanej i napisanej aplikacji w języku Java. Wyniki działania drzew decyzyjnych zebrane zostały w tabeli 2.18.



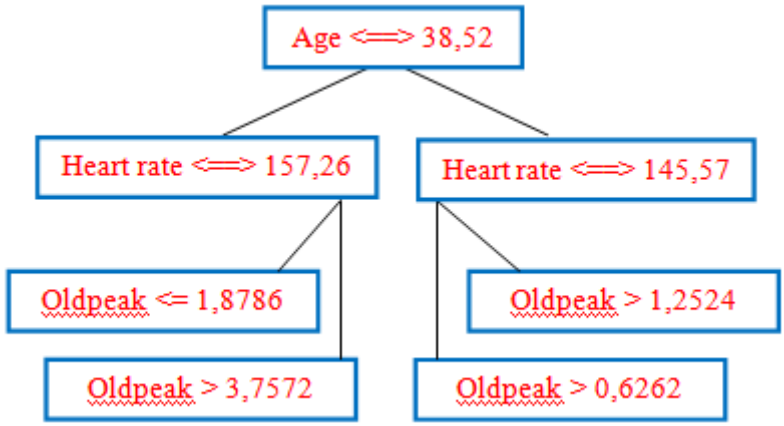
Rysunek 2.8. Zaprojektowane drzewo decyzyjne (1)

Źródło: opracowanie własne.



Rysunek 2.9. Zaprojektowane drzewo decyzyjne (2)

Źródło: opracowanie własne.



Rysunek 2.10. Zaprojektowane drzewo decyzyjne (3)

Źródło: opracowanie własne

Najwyższą skuteczność uzyskana została dla drzewa decyzyjnego z rysunku 5.10. Dla najlepszego przypadku z 47 chorych pacjentów 30 zostało poprawnie zakwalifikowanych. Zdrowi pacjenci zostali rozpoznani ze skutecznością 86% (43 z 50).

Tabela 2.18. Skuteczność modelu opartego o drzewa decyzyjne

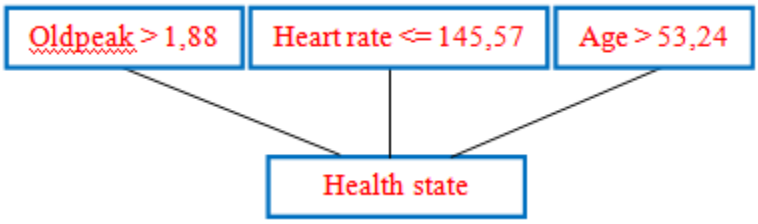
Warunek	Pozytywna prawda	Negatywna prawda	Pozytywny fałsz	Negatywny fałsz	Skuteczność modelu (%)
Age, Cholesterol (rys. 2.8)	20/47	33/50	17/50	27/47	54.64
Age, Heart rate (rys. 2.9)	32/47	27/50	23/50	15/47	60.82
Age, Heart rate, Oldpeak (rys. 2.10)	30/47	43/50	7/50	17/47	75.26

Źródło: opracowanie własne

2.8. Sieć Bayesa

Inną z metod, jaka została zaimplementowana do klasyfikacji bazy danych pod kątem oceny stanu kardiologicznego pacjenta jest sieć Bayesa. Przy projektowaniu sieci brały udział jedynie parametry o charakterze ciągłym. Na rysunku 2.11 przedstawiona została prosta sieć Bayesa.

Wykorzystuje ona do kategoryzacji danych naiwny klasyfikator Bayesa. Jak wiadomo z przeprowadzonej analizy współzależności cech zebrane w bazie danych parametry są od siebie zależne. Może to być powodem osiągnięcia niskich wartości skuteczności zaprojektowanej sieci. W tabeli 2.19 zgromadzone zostały wyniki działania trzech prostych sieci Bayesa.



Rysunek 2.11. Zaprojektowana sieć Bayesa

Źródło: opracowanie własne.

Najlepszy wynik osiągnięty został dla sieci z rysunku 2.11 (trzeci przypadek w tabeli 2.19). Sieć ta wykazywała się bardzo dobrym rozpoznawaniem pacjentów bez dolegliwości (46/50). Zdecydowanie słabiej rozpoznawała przypadki pacjentów z zaburzeniami kardiologicznymi (17/47). Skuteczność na poziomie 65% jest najniższą z wszystkich zaproponowanych przez autora metod.

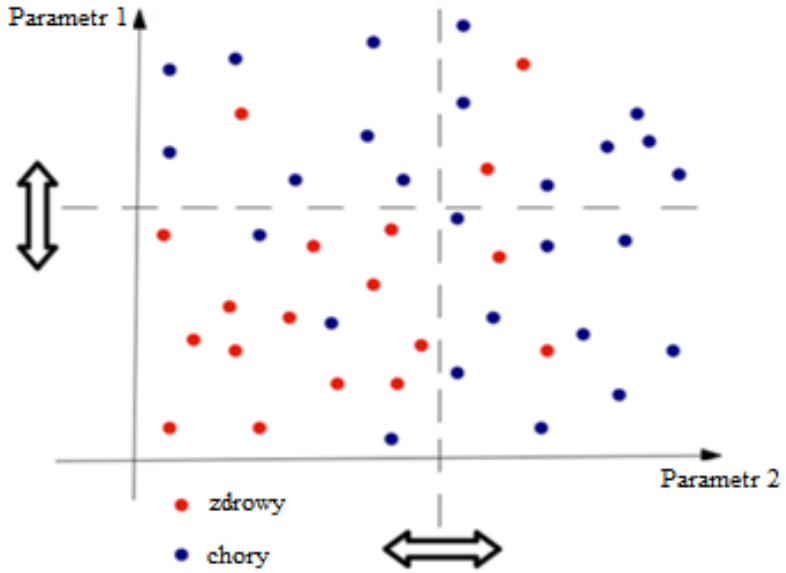
Tabela 2.19. Skuteczność modelu opartego o sieć Bayesa

Warunek	**Pozytywna prawda**	**Negatywna prawda**	**Pozytywny fałsz**	**Negatywny fałsz**	**Skuteczność modelu (%)**
(Age > 53.24), (Heart rate ≤ 145.5)	33/47	25/50	25/50	14/47	59.79
(Age > 53.24), (Cholesterol > 258)	33/47	25/50	25/50	14/47	59.79
(Age > 53.24), (Oldpeak > 1.88), (Heart rate ≤ 145.5)	17/47	46/50	4/50	30/47	64.95

Źródło: opracowanie własne

2.9. Liniowy model decyzyjny

Ostatnią powszechnie stosowaną i wykorzystaną w ramach badań metodą była klasyfikacja na podstawie wartości jednego lub kilku parametrów. W przypadku zastosowania kilku parametrów warunki łączone są za pomocą iloczynu logicznego. Geometrycznie, odpowiada to cięciom w obszarze przestrzeni wyszukiwania, która charakteryzuje się zwiększonym ryzykiem choroby (rysunek 2.12).



Rysunek 2.12. Model stratyfikacji ryzyka na podstawie iloczynu logicznego

Źródło: opracowanie własne.

Tabela 2.20. Skuteczność modelu opartego o iloczyn logiczny

Warunek	**Pozytywna prawda**	**Negatywna prawda**	**Pozytywny fałsz**	**Negatywny fałsz**	**Skuteczność modelu (%)**
Thalach ≤ 145.5	31/47	36/50	14/50	16/47	69.07
Age > 53.24 && Oldpeak > 0.6	23/47	42/50	8/50	24/47	67.01
Age > 28.71 && Oldpeak > 0.63 && Thalach ≤ 157.2	29/47	45/50	5/50	18/47	76.29

Źródło: opracowanie własne.

W tabeli 2.20 przedstawiona została skuteczność modeli opartych o iloczyn logiczny. Dla ostatniego warunku przedstawionego w tabeli skuteczność modelu wyniosła 76.29%. Nie wliczając sieci neuronowej jest to najwyższy wynik z wszystkich modeli prezentowanych w niniejszej pracy. Podobnie jak w przypadku sieci Bayesa i drzewa logicznego zbiór uczący składał się z 66% wszystkich przypadków zawartych w bazie danych.

3. PODSUMOWANIE

W monografii przedstawiono metody sztucznej inteligencji wykorzystywane przy zadaniu klasyfikacji. Algorytmy sieci neuronowych stosowane w badaniach przeprowadzonych w ramach niniejszej pracy mogą zostać zaimplementowane w docelowym systemie mikroprocesorowym, będącym przystawką do monitora funkcji życiowych.

Zaproponowany sposób redukcji pozwala na wykonanie analizy klasyfikacji i podjęcia decyzji w oparciu o mniejszą liczbę cech. Jest to kwestią istotną w przypadku takiego rozwiązania, jak omawiana przystawka do monitora funkcji życiowych, która musi analizować wiele zmiennych w trybie rzeczywistym i na ich podstawie podjąć decyzję co do sugerowanego rozpoznania lekarskiego. Układ rozpoznający, którego składowym zadaniem jest klasyfikacja, może dopuszczać zwiększenie liczby błędnych decyzji – ważniejszy w takich systemach jest limit czasu nałożony na wykonanie całości zadania. Można do zilustrowania takiego zagadnienia posłużyć się przykładem rozpoznawania twarzy w miejscach publicznych. Nie jest najważniejszym warunkiem fakt, że algorytm decyzyjny na 100% ustalił, że dana osoba jest poszukiwana, jeżeli czas potrzebny do uzyskania tego wyniku był na tyle długi, że osoba ta zdążyła dawno opuścić dane miejsce.

W przypadku danych medycznych bardzo istotne jest zachowanie odpowiedniej skuteczności decyzyjnej – błąd w takim przypadku może wiązać się z poważnymi konsekwencjami zdrowotnymi. Oczywiście nad każdym systemem stoi jeszcze pracownik medyczny, który podejmuje ostateczną decyzje co do dalszego postępowania. Nie zmienia to jednak faktu, że od systemów decyzyjnych dotyczących diagnozy i wykrywania sytuacji krytycznych wymaga się szczególnie wysokiej jakości predykcji.

Powyższe rozważania wskazują, że wspomniana minimalizacja danych wejściowych (szybkość działania modułu decyzyjnego) ma w przypadkach medycznych duże znaczenie, ale nie mniejsze niż dokładności przewidywania algorytmów sztucznej inteligencji.

W celu weryfikacji zaproponowanej metody przeprowadzony został szereg czasochłonnych eksperymentów, które zostały przedstawione w pracy. Uzyskane wyniki sugerują przydatność zaproponowanej metody. Rezultaty zostały otrzymane dla wielu różnych metod uczenia sieci neuronowej i różnych struktur sieci. Każdy z przypadków podanych w tabelach wyników został uśredniony ze 100 prób uczenia. Daje to możliwość generalizacji

wyników, a nie porównania jedynie najlepszych przypadków uczenia sieci (w przypadku, gdy wagi początkowe zostały optymalnie dobrane).

Przeprowadzone badania udowodniły, że dekompozycja parametrów jakościowych stanu kardiologicznego pacjenta wraz z usunięciem liniowych zależności wszystkich parametrów umożliwia poprawę skuteczności rozpoznania stanu kardiologicznego przy jednoczesnym zachowaniu tak samo licznego zbioru wejściowego.

Na podstawie zrealizowanych eksperymentów można zauważyć, że konstrukcja nowych niezależnych liniowo cech w modelu stratyfikacji ryzyka zgonu umożliwia znaczną redukcję liczby danych wejściowych pacjenta bez utraty użytecznej informacji predykcyjnej i przy zachowaniu jakości klasyfikacji uzyskiwanej z zastosowaniem pierwotnego zbioru parametrów fizjologicznych pacjenta.

Opisane w monografii prace badawcze obejmowały opracowanie i zaimplementowanie metod redukcji zbioru różnego rodzaju parametrów fizjologicznych pacjenta, a także weryfikację opracowanych algorytmów przy użyciu rzeczywistych zbiorów danych.

BIBLIOGRAFIA

[1] Wojtyniak B. i in., *Umieralność przedwczesna i ogólna z powodu chorób układu krążenia w Polsce na tle sytuacji w Unii Europejskiej, Europie i USA*. Warszawa: Państwowy Zakład Higieny, 2007.

[2] Wojtysiak B. i in., *Sytuacja zdrowotna ludności Polski*. Warszawa: Narodowy Instytut Zdrowia Publicznego – Państwowy Zakład Higieny, 2008.

[3] Cooper S., Cade J., *Predicting survival, in-hospital cardiac arrest: Resuscitation survival variables and training effectiveness*. "Resuscitation" 1997, Vol. 35 (1), s. 17-22.

[4] Katarzyński P. i in., *Design of elliptic filters with phase correction by using genetic algorithm*, "Przegląd Elektrotechniczny" 2010, R. 86, nr 11A, s. 69-73.

[5] Katarzyński P. i in., *Genetic algorithms in gyrator-capacitor filters,* [w:] *Proceedings of the 18th International Conference - Mixed Design of Integrated Circuits and Systems, MIXDES 2011*, 2011, s. 602-607.

[6] Subbe C.P., Kruger M., Gemmel L., *Validation of a modified Early Warning Score in medical admissions*. "Quarterly Journal of Medicine" 2001, nr 94, s. 521-526.

[7] McNelis J. i in., *A comparison of predictive outcomes of APACHE II and SAPS II in surgical intensive care unit*. "American Journal of Medical Quality" 2001, nr 16, s. 161-165.

[8] Knaus W.A. i in., *The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults*. "Chest." 1991, Vol. 100(6), s. 1619–1636.

[9] Miranda D.R., de Rijka A., Schaufeli W., *Simplified Therapeutic Intervention Scoring System: The TISS-28 items. Results from a multicenter study*. "Critical Care Medicine" 1996, vol. 24, s. 643.

[10] Tadeusiewicz R., *Sieci neuronowe*. Warszawa: Akademicka Oficyna Wydawnicza RM, 1993.

[11] Wawryn K., Strzeszewski B., *Current mode circuits for programmable WTA neural network*, "Analog Integrated Circuits and Signal Processing" 2001, Vol. 27, iss. 1-2, s. 49-69.

[12] Wawryn K., *AB class current mode multipliers for programmable neural networks*, "Electronics Letters" 1996, Vol. 32, iss. 20, s. 1902-1904.

[13] Wemmenhove B. i in., *Inference in the Promedas Medical Expert System*. [w:] *Artificial Intelligence in Medicine. AIME 2007*. Bellazzi R., Abu-Hanna A., Hunter J. (eds), Springer, Berlin, Heidelberg, 2007, s. 456-460 [Lecture Notes in Computer Science, vol. 4594].

[14] Wiggins M. i in., *Evolving a Bayesian classifier for ECG-based age classification in medical applications*. "Applied Soft Computing" 2008, vol. 8, s. 599-608.

[15] Laskey K.B., *MEBN: A Language for First-Order Bayesian Knowledge Bases*. "Artificial Intelligence" 2008, vol. 172, s. 140-178.

[16] Khemphila A., Boonjing V., *Comparing performances of logistic regression, decision trees, and neural networks*. [w:] *2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, 2010, pp. 193-198.

[17] Tylman W. i in., *System for estimation of patient's state – discussion of the approach*, "New Results in Dependability and Computer Systems" 2013, Vol. 224, s. 501-511.

[18] Jambu M., *Exploratory and multivariate data analysis*. Orlando: Academic Press, 1991.

[19] Khemphila A., Boonjing V., *Heart disease Classification using Neural Network and Feature Selection*. [w:] *2011 21st International Conference on Systems Engineering*, 2011, s. 406-409.

[20] de Leeuw J., *Principal Component Analysis of Binary Data by Iterated Singular Value Decomposition*, Department of Statistics, University of California Los Angeles, 2004

[21] de Leeuw J., *Principal Component Analysis of Binary Data. Applications to Roll-Call-Analysis*. Department of Statistics, UCLA, 2003.

[22] Kolenikov S., Angeles G., *The use of discrete data in PCA: Theory, simulations, and applications to socioeconomic indices*, Chapel Hill: Carolina Population Center MEASURE, 2004.

[23] Górecki T., *Podstawy statystyki z przykładami w R*. Legionowo: Wydawnictwo BTC, 2011.

[24] StatSoft, *Statistica for Windows*. Kraków: Statsoft, 1997.

[25] Spearman C., *General intelligence, objectively determined and measured*. "The American Journal of Psychology" 1904, Vol. 15, No. 2, s. 201-292.

[26] Cook D., Swyane D.F., *Interactive and Dynamic Graphics for Dara Analysis With R and GGobi*. New York: Springer, 2007.

[27] McNelis J. i in., *A comparison of predictive outcomes of APACHE II and SAPS II in surgical intensive care unit.* "American Journal of Medical Quality" 2001, vol. 16, s. 161-165.

[28] Vargas F. i in. *Electrocardiogram pattern recognition by means of MLP network and PCA: a case study on equal amount of input signal types.* [w:] *VII Brazilian Symposium on Neural Networks, 2002. SBRN 2002. Proceedings.,* 2002, s. 200-205.

[29] Hao Z., Li-Qing Z., *ECG analysis based on PCA and Support Vector Machines.* [w:] *2005 International Conference on Neural Networks and Brain*, 2005, s. 743-747.

SPIS RYSUNKÓW

SPIS TABEL