



UNIWERSYTET MEDYCZNY
IM. KAROLA MARCINKOWSKIEGO W POZNANIU
Katedra i Zakład Informatyki i Statystyki

Izabela Miechowicz

**Wybrane wielowymiarowe metody analizy
zmienności rytmu serca u pacjentów z przebyłym
zawałem mięśnia sercowego, nadciśnieniem i/lub
cukrzycą**

rozprawa na stopień doktora nauk medycznych

promotor: prof. dr hab. n. med. Jerzy A. Moczko

Poznań, 2015

PODZIĘKOWANIA

Składam serdeczne podziękowania mojemu Promotorowi prof. dr hab. Jerzemu Moczko, za wsparcie, cenne wskazówki i wszelką okazaną pomoc.

Pragnę podziękować prof. dr hab. Tomaszowi Siminiakowi wraz z Zespołem za udostępnienie danych, w oparciu o które powstała niniejsza rozprawa doktorska.

Wojciechowi dziękuję za ogromne wsparcie i bezgraniczną wiarę w to, że się uda.

Serdecznie dziękuję Mamie i Tacie za to, że zawsze mogłam na nich liczyć.

Spis treści

1. WYKAZ SKRÓTÓW	7
2. WSTĘP.....	10
2.1 Sposoby uzyskania informacji o zmienności rytmu serca	12
2.1.1 Pomiary w dziedzinie czasu	13
a) metody statystyczne	13
b) metody geometryczne.....	13
2.1.2 Analiza w dziedzinie częstotliwości.....	14
2.1.3 Pomiary w połączonych dziedzinach czasu i częstotliwości.....	16
2.1.4 Nieliniowe metody pomiaru zmienności rytmu serca	18
a) współczynnik nachylenia dla prawa potęgowego rozkładu mocy sygnału	18
b) beztrendowy współczynnik potęgowy skalowania fraktalnego	18
c) wykres Poincaré.....	19
d) turbulencja rytmu serca (Heart Rate Turbulence, HRT)	20
2.2 Wpływ wybranych czynników na zmienność rytmu serca.....	21
2.2.1 Choroba niedokrwienna a zmienność rytmu serca	21
2.2.2 Nadciśnienie a zmienność rytmu serca.....	24
2.2.3 Wpływ cukrzycy na zmienność rytmu serca	27
2.3 Wskaźniki zmienności rytmu serca	28
2.3.1 W dziedzinie czasu	28
2.3.2 W dziedzinie częstotliwości	35
2.3.3 W połączonych dziedzinach czasu i częstotliwości.....	36
2.3.4 Normalizacja wskaźników zmienności rytmu serca.....	38
2.3.5 Akwizycja danych (sygnałów biomedycznych)	39
3. CEL PRACY	41
4. ANALIZOWANE GRUPY	42
4.1 Grupa badana.....	42
4.1.1 Pacjenci po przebytych zawale mięśnia sercowego.....	43
4.1.2 Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie	44
4.1.3 Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie i cukrzycę.....	45
4.1.4 Pacjenci chorujący na nadciśnienie i cukrzycę.....	46
4.2 Grupa kontrolna.....	47
5. METODY.....	49

5.1	Metody jednowymiarowe	49
5.2	Metody wielowymiarowe	50
5.2.1	Regresja logistyczna	50
a)	postać funkcji logistycznej	51
b)	model logistyczny	52
c)	metody estymowania parametrów	53
d)	testowanie hipotez w modelu regresji logistycznej	54
e)	iloraz szans	56
f)	dobroć dopasowania modelu regresji logistycznej	59
5.2.2	Drzewa klasyfikacyjne	61
a)	definicja drzewa klasyfikacyjnego	61
b)	metody konstrukcji drzew	63
	<input type="checkbox"/> określenie kryterium trafności przewidywania	64
	<input type="checkbox"/> wybór metody „optymalnego” podziału węzłów	65
	<input type="checkbox"/> ustalenie reguły zatrzymania podziału i reguła decyzyjna	67
	<input type="checkbox"/> wybór właściwej wielkości drzewa	68
c)	najczęściej stosowane algorytmy w tworzeniu drzew klasyfikacyjnych	71
6.	ZASTOSOWANE OPROGRAMOWANIE	72
7.	WYNIKI	73
7.1	Analiza jednowymiarowa	73
7.1.1	Pacjenci po przebytych zawale mięśnia sercowego vs pacjenci z grupy kontrolnej ...	73
a)	kobiety	78
b)	mężczyźni	80
7.1.2	Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie vs pacjenci z grupy kontrolnej	82
a)	kobiety	85
b)	mężczyźni	85
7.1.3	Pacjenci po przebytych zawale mięśnia sercowego, którzy chorują na nadciśnienie i cukrzycę vs pacjenci z grupy kontrolnej	87
a)	kobiety	93
b)	mężczyźni	95
7.1.4	Pacjenci, którzy chorują na nadciśnienie i cukrzycę vs pacjenci z grupy kontrolnej ..	97
a)	kobiety	103
b)	mężczyźni	105

7.2 Regresja logistyczna	107
7.2.1 Pacjenci po przebytych zawale mięśnia sercowego	107
7.2.2 Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie	113
7.2.3 Pacjenci po przebytych zawale mięśnia sercowego, którzy chorują na nadciśnienie i cukrzycę	118
7.2.4 Pacjenci, którzy chorują na nadciśnienie i cukrzycę	121
7.2.5 Podsumowanie modeli regresji logistycznej	126
7.3 Drzewa klasyfikacyjne	127
7.3.1 Pacjenci po przebytych zawale mięśnia sercowego	127
7.3.2 Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie	131
7.3.3 Pacjenci po przebytych zawale mięśnia sercowego, chorujący na nadciśnienie i cukrzycę	135
7.3.4 Pacjenci chorujący na nadciśnienie i cukrzycę	138
7.3.5 Podsumowanie modeli drzew klasyfikacyjnych	141
7.4 Podsumowanie modeli	142
7.4.1 Osoby po przebytych zawale mięśnia sercowego	143
7.4.2 Osoby po przebytych zawale mięśnia sercowego i chorujące na nadciśnienie	145
7.4.3 Osoby po przebytych zawale mięśnia sercowego oraz chorujące na nadciśnienie i cukrzycę	147
7.4.4 Osoby chorujące na nadciśnienie tętnicze i cukrzycę	149
8. DYSKUSJA	151
9. WNIOSKI	165
10. STRESZCZENIE	167
11. SUMMARY	169
12. BIBLIOGRAFIA	170
ZAŁĄCZNIK 1 TABELE PRZEDSTAWIAJĄCE WYSTĄPIENIE ISTOTNYCH STATYSTYCZNIE RÓŻNIC W WARTOŚCIACH PARAMETRÓW POMIĘDZY OSOBAMI Z GRUPY BADANEJ I KONTROLNEJ	188
1. pacjenci po przebytych zawale mięśnia sercowego	188
2. pacjenci po przebytych zawale mięśnia sercowego chorujący na nadciśnienie	189
3. pacjenci po przebytych zawale mięśnia sercowego chorujący na nadciśnienie tętnicze i cukrzycę	190
4. pacjenci chorujący na nadciśnienie tętnicze i cukrzycę	191

ZAŁĄCZNIK 2 BUDOWA ANALIZOWANYCH MODELI LOGISTYCZNYCH ORAZ TABELE KORELACJI MIĘDZY ZMIENNYMI UWZGLĘDNIONYMI W MODELU	192
1. pacjenci po przebytych zawale mięśnia sercowego	192
2. pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie	196
3. pacjenci po przebytych zawale mięśnia sercowego, chorujący na nadciśnienie tętnicze i cukrzycę.....	198
4. pacjenci chorujący na nadciśnienie tętnicze i cukrzycę	199
ZAŁĄCZNIK 3 STRUKTURA DRZEW KLASYFIKACYJNYCH	204
1. Pacjenci po przebytych zawale mięśnia sercowego.....	204
2. Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie	205
3. Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie tętnicze i cukrzycę.....	206
4. Pacjenci chorujący na nadciśnienie tętnicze i cukrzycę	206
ZAŁĄCZNIK NR 4 STRUKTURA ORAZ MACIERZE BŁĘDNYCH KLASYFIKACJI DLA DRZEW INTERAKCYJNYCH	208
1. Pacjenci po przebytych zawale mięśnia sercowego.....	208
2. Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie	209
3. Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie tętnicze i cukrzycę.....	210
4. Pacjenci chorujący na nadciśnienie tętnicze i cukrzycę	211

1. Wykaz skrótów

AHA	Amerykańskie Towarzystwo Kardiologiczne (ang. AMERICAN HEART ASSOCIATION)
ASDNN	średnia odchyłeń standardowych z wszystkich interwałów NN dla wszystkich 5 minutowych fragmentów w ciągu 24h (ang. Mean of the standard deviations of all NN intervals for all 5-minute segments in 24 hours)
AUN	autonomiczny układ nerwowy (ang. Autonomic Nervous System)
BAND	wskaźnik zmienności długoterminowej Organa (ang. Organ Band index)
BMI	wskaźnik masy ciała (ang. Body Mass Index)
CAN	cukrzycowa neuropatia autonomiczna (ang. Cardiac Autonomic Neuropaty)
ChNS	Choroba Niedokrwienna Serca
CI	przedział ufności (ang. Confidence Interval)
CSD	rozkład stożkowy (ang. Cone Shaped Distribution)
Dalton MABB	wskaźnik zmienności krótkoterminowej Daltona
Dalton SD	wskaźnik zmienności długoterminowej Daltona
De Haan LTI	wskaźnik zmienności długoterminowej de Haana (ang. Long Term Irregularity index)
De Haan STI	wskaźnik zmienności krótkoterminowej de Haana (ang. Short Term Irregularity index)
DFA	beztrendowa analiza fluktuacji (ang. Detrended Fluctuation Analysis)
DI	wskaźnik różnic (ang. Differential Index)
EKG	elektrokardiogram
HF	zakres wysokich częstotliwości (0.15 Hz, 0.4 Hz> (ang. High Frequency)
HR	częstość rytmu serca (ang. Heart Rate)
HRT	Turbulencja Rytmu Serca (ang. Heart Rate Turbulence)
HRV	zmienność rytmu serca (ang. Heart Rate Variability)
Huey LTV	wskaźnik zmienności długoterminowej Hueya (ang. Long Term Variability index)

Huey STV	wskaźnik zmienności krótkoterminowej Hueya (ang. Short Term Variability index)
ID	różnice odstępów (ang. Interval Differences)
II	wskaźnik odstępów (ang. Interval Index)
IQR	rozstęp międzykwartylowy (ang. Interquartile Range)
JTFA	analiza w połączonych dziedzinach czasu i częstotliwości (ang. Joint Time-Frequency Analysis)
LF	zakres niskich częstotliwości (0.04Hz, 0.15 Hz> (ang. Low Frequency)
LR	regresja logistyczna (ang. Logistic Regression)
LTI	wskaźnik zmienności długoterminowej (ang. Long Term Irregularity index)
LTV	wskaźnik zmienności długoterminowej (ang. Long Term Variability index)
max	wartość maksymalna
min	wartość minimalna
OR	iloraz szans (ang. Odds ratio)
OSC	szerokość wstęgi oscylacyjnej (ang. Oscillation Index)
p	wartość prawdopodobieństwa
P(A)	prawdopodobieństwo wystąpienia zdarzenia A
PCI	przezskórna interwencja wieńcowa (ang. Percutaneous Coronary Interventions)
$P(Y=1/x_1, x_2, \dots, x_n)$	model regresji logistycznej
Q ₁	kwartył dolny
Q ₃	kwartył górny
RMSSD	wartość średnia kwadratowa z kolejnych różnic (ang. Root Mean Square of Successive Differences)
SA	węzeł zatokowo-przedsionkowy (ang. Sino-atrial)
S(A)	szansa wystąpienia zdarzenia A
SCD	nagły zgon z przyczyn naczyniowo - sercowych (ang. Sudden Cardiac Death)
SD	odchylenie standardowe (ang. Standard Deviation)
SDANN	odchylenie standardowe z 5-ciominutowych średnich interwałów NN (ang. Standard Deviation of 5-minute average NN intervals)

SDNN	odchylenie standardowe czasów trwania wszystkich zarejestrowanych odstępów RR rytmu zatokowego (ang. Standard Deviation of NN)
STI	wskaźnik zmienności krótkoterminowej (ang. Short Term Irregularity index)
STV	wskaźnik zmienności krótkoterminowej (ang. Short Term Variability index)
TIMI	określenie przepływu wieńcowego (ang. Thrombolysis In Myocardial Infarction)
TINN	interpolacja trójkątna histogramu interwałów NN (ang. Triangular Interpolation of NN interval histogram)
TO	początek turbulencji (ang. Turbulence Onset)
TS	nachylenie turbulencji (ang. Turbulence Slope)
van Geijn ID	wskaźnik zmienności krótkoterminowej van Geijna
\bar{x}	średnia arytmetyczna
x_1, x_2, \dots, x_n	zmienne niezależne, predyktory
Y	dychotomiczna zmienna zależna
Yeh DI	wskaźnik zmienności krótkoterminowej Yeh'a
Yeh II	wskaźnik zmienności długoterminowej Yeh'a
Zugaib LTV	wskaźnik zmienności długoterminowej Zugaiba
Zugaib STV	wskaźnik zmienności krótkoterminowej Zugaiba

2. Wstęp

Zmienność rytmu serca (Heart Rate Variability, HRV), która jest generowana przez aktywność węzła zatokowego-przedsionkowego (sino-atrial, SA), odzwierciedla działania autonomicznego układu nerwowego (Autonomic Nervous System, AUN), zwłaszcza jego części przywspółczulnej[1]. Komórki węzła SA stale generują impuls elektryczny, który inicjuje proces skurczu mięśnia sercowego (zweżanie kolejnych 4 komór serca: 2 komór i 2 przedsionków). W stanie spoczynku węzeł SA generuje 100-120 impulsów na minutę. Jednakże u zdrowego człowieka w stanie spoczynku tętno wynosi przeciętnie 55-70 uderzeń na minutę, co wynika z tego, że część impulsów jest hamowanych przez autonomiczny układ nerwowy, który sprawuje kontrolę nad pracą węzła zatokowo-przedsionkowego.

Autonomiczny układ nerwowy to część układu nerwowego nie będąca pod świadomym wpływem. Jego funkcją jest sprawowanie kontroli nad pracą narządów i układów w organizmie, które pracują rytmicznie, regularnie i automatycznie, np. nad oddychaniem, częstością akcji serca czy trawieniem. AUN składa się z dwóch składowych: współczulnej i przywspółczulnej.

Część współczulna AUN odpowiada za zmobilizowanie organizmu do „walki” w momencie zagrożenia, jakim może być np. stres, lub w innych sytuacjach wymagających samoobrony. Wzrost pobudzenia współczulnego powoduje wzrost częstości akcji serca, jednakże jest on dość powolny, gdyż zwiększenie częstości rytmu serca (Heart Rate, HR) można zauważyć dopiero po upływie 5 s od rozpoczęcia stymulacji współczulnej, aż do 30 s, kiedy to osiąga najwyższy poziom.

Część przywspółczulna AUN jest przeciwieństwem w stosunku do układu współczulnego. Podczas stymulacji przywspółczulnej spada tętno, ruszają procesy trawienne oraz zaczynają działać mechanizmy odpowiedzialne za naprawę tkanek i narządów. HR bardzo szybko reaguje na stymulację przywspółczulną, spowolnienie rytmu serca jest niemal natychmiastowe i przebiega w ciągu 1-2 uderzeń serca [2].

U osób zdrowych rytm serca jest zmienny. Większa zmienność HR koreluje z lepszym zdrowiem, gdyż wskazuje na istnienie równowagi pomiędzy częścią współczulną i przywspółczulną AUN. Im niższe wartości HRV, tym większy brak równowagi w funkcjonowaniu autonomicznego układu nerwowego, i tym większe ryzyko wystąpienia choroby obecnie i w przyszłości. Dlatego też analizę HRV stosuje się jako narzędzie do badań przesiewowych, gdyż zmiany w nim poprzedzają zmiany pojawiające się w wielu procesach chorobowych. Wykazano, że wahania w HRV mogą poprzedzać objawy kliniczne pewnych chorób neurologicznych. Ze względu na prostotę i nieinwazyjność metod

pozwalających na ocenę HRV, badania dotyczące zmienności rytmu serca obejmują wiele dziedzin medycyny. Starano się wykorzystać jej wartości kliniczne i prognostyczne w badaniach dotyczących cukrzycy oraz chorób neurologicznych takich jak np. udar mózgu, choroba Parkinsona czy padaczka. Udowodniono również jej rolę jako markera ryzyka naczyniowo-sercowego.

Zmienność rytmu serca jest ważnym wskaźnikiem działania autonomicznych funkcji układu krążenia [3]. Wg klasycznej teorii homeostazy każdy układ dąży do zachowania stałości swoich parametrów. W przypadku układu sercowo-naczyniowego dążenie to widoczne jest w utrzymywaniu względnie stałych wartości rytmu serca oraz ciśnienia tętniczego. Jednakże, ze względu na specyfikę przebiegu procesów fizjologicznych, mimo zapewnienia możliwie stałych warunków środowiskowych i braku bodźców zewnętrznych mogących powodować zakłócenia, można zauważyć fluktuacje rytmu serca oraz ciśnienia tętniczego.

Parametry zmienności rytmu serca mierzone w dziedzinie czasu, częstotliwości i w połączonych dziedzinach czasu i częstotliwości obrazują jak przebiegają zmiany napięcia neurowegetatywnego.

Związek pomiędzy zmiennością rytmu serca a oddychaniem opisał w 1733 r. Stephen Hales [4], jeden z wynalazców wentylatora. Zastosowanie tego wynalazku poprawiło wymianę powietrza w szpitalach lub statkach, dzięki czemu zwiększyło prawdopodobieństwo przeżycia pacjentów. W 1965 roku Hon i Lee [5] dostrzegli związek pomiędzy obniżoną zmiennością rytmu serca płodu podczas porodu a zagrożeniem jego życia, zanim jeszcze zaobserwowano zmiany w częstości akcji serca. Kolejne badania dowiodły wartości diagnostycznej parametrów HRV w przypadku neuropatii cukrzycowej [6,7].

W 1978 r. Wolf i wsp. [8] zaobserwowali u pacjentów z pozawałową dysfunkcją lewej komory obniżoną wartość zmienności rytmu serca i wiązali z tym niekorzystne rokowania. Badania prowadzone przez Kleigera i wsp. [9] w latach 80 XX w. potwierdzają tezę o wartości prognostycznej wskaźników HRV u pacjentów po przebytych zawale mięśnia sercowego oraz wskazują na silny związek pomiędzy obniżonymi wartościami HRV a zwiększoną śmiertelnością w ciągu pierwszych trzech lat po ostrym zawale. Bigger i wsp. [10] wykazali, że znaczące obniżenie aktywności układu przywspółczulnego u chorych po zawale mięśnia sercowego może być czynnikiem ryzyka śmierci. Badania nad wykorzystaniem współczynników HRV w celach diagnostycznych bardzo przybrały na sile pod koniec XX wieku, gdyż liczone na to, iż będzie można stosować w praktyce klinicznej analizę HRV jako nieinwazyjne i stosunkowo niedrogie narzędzie.

Celem współczesnej kardiologii jest zmniejszenie śmiertelności związanej z chorobami układu sercowo-naczyniowego. W obecnych czasach ma to ogromne znaczenie, gdyż szacuje się, że w Europie choroby układu krążenia są najczęstszą przyczyną zgonów, a śmiertelność z powodu schorzeń sercowo-naczyniowych, w tym zawału mięśnia sercowego, jest dwa razy większa niż z powodu chorób nowotworowych [11]. Dlatego też badania nad nowymi modelami wykorzystującymi zmienność rytmu serca trwają nadal [12]. Białek-Ławniczak i wsp. [13] oraz Kuzemczak i wsp. [14] dowodzą, że analiza częstotliwościowa oraz analiza w połączonych dziedzinach czasu i częstotliwości zmienności rytmu serca u pacjentów chorujących na stabilną chorobę wieńcową może pomóc w identyfikacji osób wymagających pilnego zabiegu angioplastyki wieńcowej (percutaneous coronary interventions, PCI). Wykazano również u tych pacjentów istnienie wyższych wartości szerokości wstęgi oscylacyjnej (oscillation index, OSC) w przypadku, gdy przeszli oni udar mózgu w ciągu 5 lat od badania [15].

Nasilenie zmian miażdżycowych u pacjentów chorujących na stabilną chorobę wieńcową, mierzone w skali Gensiniego, znajduje swoje odzwierciedlenie w obniżonej zmienności rytmu serca. Wykazano istnienie ujemnej korelacji pomiędzy wartościami wskaźników zmienności rytmu serca a punktacją uzyskaną w skali Gensiniego [16].

Do tej pory, z powodu niskiej czułości i swoistości metod używanych do analiz zmienności rytmu serca (co obniżało z kolei wartości diagnostyczne tych analiz), parametry zmienności rytmu serca były w niewielkim stopniu wykorzystywane w praktyce klinicznej. Analiza zmienności rytmu serca ma tą przewagę nad innymi nieinwazyjnymi metodami (np. testem wysiłkowym), że zmiany w HRV mogą pojawiać się wcześniej niż objawy kliniczne. Ciągły postęp związany z cyfrowym przetwarzaniem sygnałów, rozwój w dziedzinie nauk biomedycznych oraz stosowanie coraz bardziej zaawansowanych metod statystycznych daje możliwość coraz dokładniejszej, nieinwazyjnej oceny zjawiska zmienności, co może odegrać znaczącą rolę w diagnostyce kardiologicznej w procesie identyfikacji pacjentów znajdujących się w grupie podwyższonego ryzyka. Pozwoli to określić dalszy kierunek badań oraz skrócić czas potrzebny do zastosowania odpowiedniego leczenia.

2.1 Sposoby uzyskania informacji o zmienności rytmu serca

Bazą dla pomiarów HRV jest sygnał elektrokardiograficzny (obrazujący czynność bioelektryczną komórek mięśnia sercowego), który jest zaliczany do sygnałów ciągłych (analogowych).

2.1.1 Pomiary w dziedzinie czasu

a) metody statystyczne

Do metod analizy zmienności w dziedzinie czasu zalicza się metody statystyczne i geometryczne. Obie opierają się na analizie czasu trwania odstępów RR rytmu zatokowego bądź różnic pomiędzy nimi [17, 18, 19]. Najczęściej analizowanym parametrem jest odchylenie standardowe czasów trwania wszystkich zarejestrowanych odstępów RR rytmu zatokowego (Standard Deviation of NN, SDNN), mierzony w 24 godzinnym okresie czasu. Dokładne obliczenie parametru SDNN wymaga niezwyklej staranności, by wykluczyć uderzenia ektopowe, brakujące oraz artefakty. Dlatego też klasycznie oceniany parametr SDNN wymaga długotrwałych zapisów analizy holterowskiej (większość laboratoriów wymaga co najmniej 18-godzinnych zapisów prawidłowych danych, by obliczyć SDNN w 24 godzinnym okresie czasu) [20].

Kolejnym, bardzo często analizowanym parametrem jest wartość średnia kwadratowa z kolejnych różnic (Root Mean Square of Successive Differences, RMSSD), który opisuje zjawiska podlegające zmianom okresowym [21].

b) metody geometryczne

Cykl odstępów NN można przekształcić na model geometryczny (np. rozkład Lorenza odstępów NN lub RR). Zmienność ocenia się analizując właściwości geometryczne lub/i graficzne uzyskanego modelu. Najczęściej stosuje się trzy podstawowe metody:

- analizę podstawowych miar modelu geometrycznego (np. szerokość histogramu rozkładu na określonym poziomie);
- model geometryczny jest przybliżany do matematycznie określonego kształtu (np. aproksymacja trójkątem histogramu rozkładu); wówczas analizuje się matematyczne parametry związane z konkretnym kształtem;
- kształt matematyczny jest podzielony na kilka głównych, bazujących na modelu, kategorii, które odpowiadają różnym klasom HRV (np. eliptyczne, liniowe lub trójkątne kształty rozkładów Lorenza) [21].

Do metod geometrycznych pomiarów zmienności rytmu serca zalicza się HRV indeks trójkątny (Heart Rate Variability Triangular Index). Technika ta była szeroko stosowana przez badaczy ze Szpitala Św. Jerzego w Londynie [22, 23, 24]. Jest on całką z funkcji gęstości rozkładu, tzn. aby obliczyć HRV index należy liczbę wszystkich odstępów NN podzielić przez maksimum funkcji gęstości rozkładu.

Innym analizowanym parametrem jest interpolacja trójkątna histogramu interwałów NN (Triangular Interpolation of NN interval histogram, TINN), obliczana jako długość podstawy trójkąta aproksymującego rozkład odstępów NN (trójkąt ten jest wyznaczany przy użyciu metody najmniejszych kwadratów).

Na oba te wskaźniki, wyrażające zmienność całkowitą mierzoną w ciągu 24 h, większy wpływ mają niższe częstotliwości [25].

Metody geometryczne, ze względu na konieczność uzyskania określonej liczby możliwych do przyjęcia odstępów RR, nie mogą być stosowane do oceny zmienności krótkoterminowej. Metody geometryczne można stosować nawet w przypadku zapisów o niezbyt dobrej jakości, co jest niewątpliwie ich zaletą. Dzięki zastosowaniu tej metody minimalizuje się wpływ odstających wielkości odstępów RR (czyli tych znacznie dłuższych lub znacznie krótszych od przeciętnych), przez co znacznie ogranicza się wpływ uderzeń ektopowych, brakujących i artefaktów. Główną ich wadą jest czas obserwacji, gdyż preferowane są zapisy 24 godzinne (mimo iż wiadomo, że zapisy co najmniej 20 minutowe są już wystarczające).

2.1.2 Analiza w dziedzinie częstotliwości

Analiza zmienności w dziedzinie częstotliwości ukazuje informacje znajdujące się w sygnale elektrokardiograficznym (EKG) związane z cyklicznością zjawisk. Pozwala na rozpoznanie i ocenę wpływu zarówno układu współczulnego, jak i parasympatycznego na pracę serca. Jest to powiązane z procesami fizjologicznymi takimi jak: oddychanie, funkcja baroreceptorów, funkcja chemoreceptorów, aktywność adrenergiczna. Analiza spektralna została zaproponowana przez Akselroda i wsp. [26] w 1981 roku.

Według Sayers'a [27] zmienność rytmu serca można opisać za pomocą trzech podstawowych składowych dziedzin częstotliwości, co wynika właśnie z przesłanek fizjologicznych:

- składowa niskich częstotliwości [low – frequency component: (0.02Hz – 0.09 Hz), LF]- opisuje wpływy zarówno współczulne jak i

przywspółczulne [28, 29], wiąże się z aktywnością wazomotoryczną i termoregulacją [30]; wykazano, że wzrasta ona w czasie wysiłku umysłowego i umiarkowanego wysiłku fizycznego;

- składowa średnich częstotliwości [mid – frequency component: (0.09 Hz – 0.15 Hz), MF] – związane z wrażliwością baroreceptorów oraz regulacją ciśnienia krwi;
- składowa wysokich częstotliwości [high frequency: (0.15 Hz – 1.1 Hz), HF] – jako odzwierciedlenie aktywności nerwu błędnego (aktywności przywspółczulnej).

Zastosowane do obliczeń narzędzie wirtualne opisane w pracy Moczko [31] pozwala na wyznaczenie w dziedzinie niskich częstotliwości dodatkowych obszarów:

- składowa ultra niskich częstotliwości [Ultra-Low-Frequency range power, ULF: <DC, 0.003 Hz>, gdzie DC – częstotliwość prądu stałego (Direct Current) – wartość bardzo mała, bliska 0];
- składowa bardzo niskich częstotliwości [<0.003Hz, 0.04 Hz>]

Należy pamiętać jednak, iż wyznaczenia określonego pasma częstotliwości jest uwarunkowane długością analizowanego sygnału (im niższy zakres częstotliwości, tym dłuższy musi być badany sygnał).

Występowanie niektórych chorób powoduje zmiany wartości parametrów HRV mierzonych w dziedzinie częstotliwości, np. obniżenie wartości współczynnika HF ma związek z występowaniem neuropatii cukrzycowej [32]. Zmiany pozycji [33] oraz stres psychiczny [34] powodowały zmiany w wartościach wskaźnika LF/HF.

Wykazano istnienie korelacji pomiędzy wskaźnikami mierzonymi w dziedzinie czasu i odpowiednimi wskaźnikami zmierzonymi w dziedzinie częstotliwości [35, 21].

Tabela 1 Skorelowane parametry mierzone w dziedzinie czasu i częstotliwości; SDANN – odchylenie standardowe 5 minutowych, średnich interwałów NN, ASDNN – średnia odchyłeń standardowych z wszystkich interwałów NN dla wszystkich 5 minutowych fragmentów mierzonych w czasie 24 h [20]

Wskaźniki mierzone w dziedzinie czasu	Wskaźniki mierzone w dziedzinie częstotliwości
SDNN	TP (moc całkowita, ang. Total Power)
SDANN	ULF
ASDNN	VLF
RMSSD	HF

2.1.3 Pomiary w połączonych dziedzinach czasu i częstotliwości

Technika Joint Time-Frequency Analysis (JTFA) pozwala na jednoczesne zaobserwowanie zmian w zakresie częstotliwości badanego sygnału i jego zmienności w czasie. Złożoność obliczeniowa algorytmów stosowanych do realizacji analizy była przyczyną tego, iż dotychczas analiza w połączonych dziedzinach czasu i częstotliwości nie była powszechnie stosowana. Gwałtowny rozwój technologiczny, głównie w dziedzinie informatyki, spowodował, że wyniki analiz można uzyskać w rozsądnym czasie, co przyczyniło się do popularyzacji tej techniki i częstszych jej zastosowań.

Aby dokonać analizy charakterystyki sygnału jednocześnie w dziedzinie czasu i częstotliwości, konieczne jest zbudowanie funkcji elementarnych o następujących właściwościach:

- muszą być jednocześnie umieszczone w czasie i częstotliwości,
- powinny być prostopadłe do siebie i mieć długość jednostkową (lub jak najbardziej zbliżone do ortonormalnych),
- powinny gwarantować maksymalną rozdzielczość w obu dziedzinach,
- powinny jak najbardziej ograniczać tzw. składowe interferencyjne, które pojawiają się, gdy w widmie JTF występuje więcej niż jeden pik i odzwierciedlają korelację między odpowiadającą parą pików,
- powinny mieć możliwie najłatwiejszą interpretację fizyczną.

Istnieją różne algorytmy, które, w większym lub mniejszym stopniu, spełniają te założenia, np. algorytm Gabora, stożkowy, Wignera – Ville’a lub, nie zakładające określonego rozkładu statystycznego analizowanego sygnału, algorytmy adaptacyjne.

W metodzie opracowanej przez Dennisa Gabora sygnał jest rozwijany na ważony dyskretnymi współczynnikami szereg składowych sinusoidalnych modulowanych funkcjami Gaussa. Rozdzielczość czasowa i częstotliwościowa analizy jest określona przez użyty rząd rozwinięcia – parametr d .

Kolejnym rozkładem, który znalazł zastosowanie w analizie JTFA jest, opracowany wstępnie na potrzeby mechaniki kwantowej, rozkład Wignera – Ville’a. Zaletą tej metody jest to, iż uzyskuje się najlepsze parametry rozdzielczości czasowej i częstotliwościowej, wadą natomiast, że uzyskany spektrogram jest zaburzony silnymi członami interferencyjnymi, generowanymi przez parę istniejących składowych częstotliwościowych. Złożoność rozkładu Wignera – Ville’a rzeczywistych sygnałów generowanych przez układy biologiczne skłoniła do dalszych poszukiwań algorytmów umożliwiających analizę sygnału w połączonych dziedzinach czasu i częstotliwości.

Metoda zaproponowana przez Leona Cohena w roku 1966 polegała na tworzeniu wygładzonych rozkładów czasowo – częstotliwościowych, w oparciu o konwolucję funkcji podcałkowej rozkładu Wignera – Ville’a z jądrem funkcji wygładzającej, co w dużym stopniu niwelowało człony interferencyjne. Największą trudność stanowił tutaj dobór odpowiedniej funkcji wygładzającej, spełniającej zdefiniowane warunki brzegowe rozkładu czasowo – częstotliwościowego. Do metod zaliczanych do klasy Cohena należy m. in. rozkład stożkowy (Cone Shaped Distribution, CSD), w której do ustalenia stopnia usunięcia członów interferencyjnych stosuje się parametr α (im wyższa jest jego wartość, tym słabsze jest tłumienie członów interferencyjnych, z zatem lepsza jest rozdzielczość czasowo – częstotliwościowa).

Kolejną metodą analizy sygnału w połączonych dziedzinach czasu i częstotliwości jest metoda adaptacyjna, w której do rozkładu sygnału używa się ważonej sumy liniowych samodostosowujących się sygnałów świergotowych, modulowanych obwiednią gaussowską. Metoda ta nie generuje członów interferencyjnych i znajduje szczególne zastosowanie w analizie sygnałów zmieniających się w czasie w sposób powolny [36, 37, 38].

2.1.4 Nieliniowe metody pomiaru zmienności rytmu serca

Nieliniowe metody pomiaru zmienności rytmu serca są próbą ilościowego opisu struktury i złożoności odstępów RR. Jasne jest, że przykładowa, losowa seria odstępów RR oraz normalny cykl odstępów RR mogą mieć identyczną wartość SDNN, jednak ich struktura może różnić się znacząco od siebie.

Metody nieliniowe pozwalają na osiągnąć dokładny opis modelu, jednak należy pamiętać m. in. o konieczności budowy algorytmu w taki sposób, by był on niewrażliwy na naturalnie występujące losowe szумы. Należy również uwzględnić niestacjonarność sygnałów rzeczywistych, której występowanie jest przyczyną błędów estymacji.

Istnieje wiele różnych nieliniowych miar HRV, jednak tylko niektóre z nich okazują się mieć zastosowanie w ocenie ryzyka. Należą do nich:

- współczynnik nachylenia dla prawa potęgowego rozkładu mocy sygnału;
- beztrendowy współczynnik potęgowy skalowania fraktalnego;
- wykres Poincaré;
- turbulencja rytmu serca.

a) współczynnik nachylenia dla prawa potęgowego rozkładu mocy sygnału

W normalnym rytmie zatokowym, moc widma mierzona w 24 godzinnym oknie czasowym, pokazuje stopniowy, wykładniczy wzrost amplitudy wraz ze spadkiem częstotliwości [39]. Zależność tą można również przedstawić jako log mocy do log częstotliwości. Zastosowanie przekształcenia logarytmicznego powoduje przekształcenie krzywej wykładniczej na prostą, której nachylenie można poddać analizie. Wykazano, że wraz z obniżeniem wartości nachylenia rośnie ryzyko śmierci po zawale mięśnia sercowego [40].

b) beztrendowy współczynnik potęgowy skalowania fraktalnego

Jest to miara, ozn. też jako α_1 , która jest wyznaczana metodą beztrendowej analizy fluktuacji (Detrended Fluctuation Analysis, DFA) [41]. Zazwyczaj mierzy się ją kilkakrotnie w ciągu 1000 odstępów RR, a następnie uśrednia. Prawidłowe wartości wynoszą ok. 1,05. Obniżone wartości wskaźnika α_1 wiążą się z przebytym zawałem mięśnia sercowego [42, 43].

c) wykres Poincaré

Wykres Poincaré jest graficzną metodą przedstawienia korelacji pomiędzy kolejnymi interwałami RR, tzn. RR_{i+1} i RR_i .

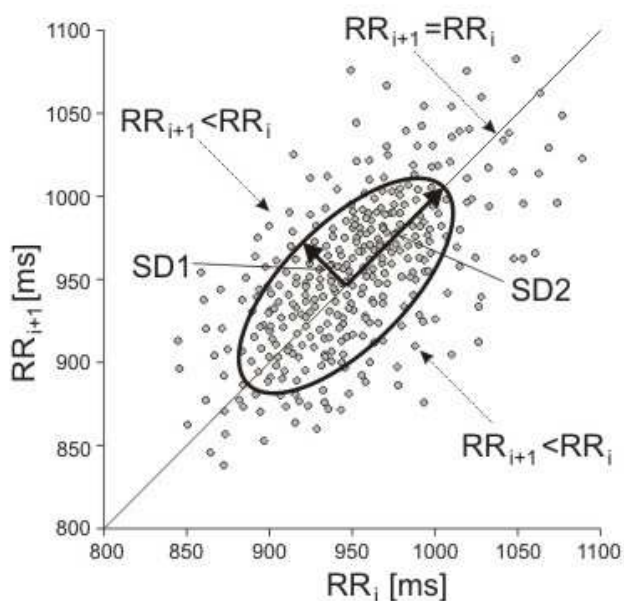
Do opisu wykresu Poincaré najczęściej służą dwa parametry:

- oś mała elipsy (SD1) – określana jest jako odchylenie standardowe mierzące rozproszenie punktów na wykresie Poincaré prostopadle do prostej $y=x$, zwanej prostą identyczności. Wszystkie punkty z wykresu Poincaré są rzutowane na prostą prostopadłą do prostej $y=x$, a następnie oblicza się odchylenie standardowe otrzymanego rozkładu [44]. Parametr ten opisuje zmienność krótkoterminową [45].
- oś wielka elipsy (SD2) – to odchylenie standardowe mierzące rozproszenie punktów na wykresie Poincaré równoległe do prostej $y=x$. Punkty w wykresie zostają rzutowane na prostą $y=x$, po czym oblicza się odchylenie standardowe. SD2 interpretowany jest jako miara zmienności zarówno krótko- jak i długoterminowej [45].

Dodatkowo analizuje się dwa kolejne parametry:

S – odpowiada polu powierzchni fikcyjnej elipsy, której półosią wielką jest SD2 (równoległa do $y=x$), a półosią małą jest SD1 (prostopadła do $y=x$). Guzik i wsp. [44] interpretują ten parametr jako miarę całkowitej zmienności rytmu serca.

SD2/SD1 – obliczany jest jako iloraz wielkości SD2 do SD1 [46, 47].



Wykres 1 Przykładowy wykres Poincaré interwałów RR, na podstawie Guzik P. i wsp. „, Correlations between the Poincaré plot and Conventional Heart Rate Variability Parameters Assessed during Paced Breathing; J. Physiol. Sci.; vol. 57, No. 1, 2007; 63-71 [44]

Wykazano istnienie zależności pomiędzy parametrami opisującymi wykres Poincaré a wskaźnikami zmienności rytmu serca mierzonymi w dziedzinie czasu i częstotliwości, np. SD1 jest idealnie skorelowany z RMSSD (z matematycznego punktu widzenia oba parametry są równoważne), co powoduje, że jest traktowany jako miara zmienności krótkoterminowej [21, 20, 48].

Względna stałość odstępów RR znajduje swoje odzwierciedlenie w kolistym kształcie wykresu Poincaré i jego niewielkiej powierzchni.

Miary te nie znajdują zastosowania w ocenie ryzyka, jednakże okazały się przydatne w identyfikacji błędów, które mogą mieć istotny wpływ na obliczenia wartości HRV.

d) turbulencja rytmu serca (Heart Rate Turbulence, HRT)

Turbulencja rytmu serca jest nową metodą analityczną. Może być wykorzystana do nieinwazyjnej oceny ryzyka wystąpienia niekorzystnych incydentów sercowo-naczyniowych. Jest określana za pomocą dwóch parametrów:

początek turbulencji (Turbulence Onset, TO) - określa procentową różnicę między częstością akcji serca tuż po przedwczesnym skurczu komorowym i częstością akcji serca przed tym skurczem. Wstępnie TO określa się dla każdego pojedynczego przedwczesnego skurczu komorowego, a następnie uśrednia się wszystkie wartości TO z analizowanego zapisu EKG. Wartości dodatnie oznaczają zwolnienie, a ujemne przyspieszenie rytmu zatokowego.

nachylenie turbulencji (Turbulence Slope, TS) - odpowiada najbardziej stromej linii regresji przeprowadzonej przez każde 5 kolejnych odstępów prawidłowego rytmu w tachogramie. Obliczenia TS opierają się na tachogramie uśrednionym ze wszystkich fragmentów rytmu zatokowego po skurczu komorowym i wyrażone w ms/odstęp RR [49]

Obniżone wartości HRT są silnie związane ze wzrostem śmiertelności po przebytych zawale mięśnia sercowego [50, 51, 52].

2.2 Wpływ wybranych czynników na zmienność rytmu serca

2.2.1 Choroba niedokrwienna a zmienność rytmu serca

Choroba niedokrwienna serca (ChNS) to choroba o zróżnicowanej etiologii. Jest ona wynikiem przewlekłego procesu niedokrwienia (niewystarczającego dostarczenia do komórek mięśnia sercowego tlenu i substancji odżywczych) [53]. Rozbieżność pomiędzy zapotrzebowaniem na nie a możliwością ich dostarczenia, pomimo wykorzystania mechanizmów autoregulacyjnych, doprowadza do niedotlenienia, które nazywane jest również niewydolnością wieńcową. Niedokrwienie prowadzi do upośledzenia czynności włókien nerwowych AUN.

Nazewnictwo tej choroby różni się w zależności od pochodzenia. Objawy kliniczne ChNS, tj. ból w klatce piersiowej i duszność, powodują, że nazywana jest dusznicą bolesną. Ze względu na jej patofizjologię określana jest jako choroba wieńcowa lub niewydolność wieńcowa.

ChNS może przebiegać zarówno w postaci ostrej jak i przewlekłej. To skutkuje dodatkowym nazewnictwem. Używa się pojęcia:

- a) stabilna – w przypadku przewlekłej choroby niedokrwiennej serca;
- b) niestabilna – odpowiadająca pojęciu ostrego zespołu wieńcowego, w której rozróżnia się:
 - zawał mięśnia sercowego;
 - zawał serca zagrażający (niestabilna choroba wieńcowa).

Zaobserwowany w ostatnich latach wzrost zachorowalności na ChNS spowodował zainteresowanie środowisk naukowych czynnikami ryzyka wystąpienia tej choroby, w celu podjęcia ukierunkowanych działań edukacyjnych mających na celu poprawę jakości życia pacjentów. Podzielono je na trzy kategorie czynników: przyczynowe, warunkowe i usposabiające, przy czym przynależność do określonej grupy zależy od wpływu konkretnego czynnika na patogenezę choroby [54]. W 1998 r. Amerykańskie Towarzystwo Kardiologiczne (American Heart Association, AHA) dokonało analizy czynników ryzyka.

Tabela 2 Kategorie czynników ryzyka choroby niedokrwiennej serca, według Amerykańskiego Towarzystwa Kardiologicznego, 1998, na podstawie „Występowanie choroby niedokrwiennej serca a praca zawodowa”; Bugajska J., Jędryka-Góral A., Konarska M., Bezpieczeństwo pracy, 4/2006, 14 – 17 [54]

CZYNNIKI PRZYCZYNOWE
<ul style="list-style-type: none">• Palenie papierosów;• Nadciśnienie tętnicze;• Podwyższone stężenia cholesterolu całkowitego i cholesterolu frakcji LDL;• Cukrzyca;• Niskie stężenie cholesterolu frakcji HDL;
CZYNNIKI WARUNKOWE
<ul style="list-style-type: none">• Podwyższone stężenie trójglicerydów;• Małe gęste cząsteczki LDL;• Nieprawidłowe stężenie lipoproteiny;• Nieprawidłowe stężenie homocysteiny;• Nieprawidłowe stężenia czynników krzepnięcia;• CRP;

CZYNNIKI USPOSABJAJĄCE

- Otyłość;
- Brak aktywności fizycznej;
- Płeć męska;
- Przedwczesna choroba niedokrwienna serca w wywiadzie rodzinnym*;
- Czynniki socjoekonomiczne, etniczne i behawioralne;
- Insulinooporność.

*- wystąpienie zawału mięśnia sercowego lub nagłego zgonu sercowego u ojca lub innego krewnego I stopnia płci męskiej w wieku poniżej 55 lat lub u matki bądź krewniej I stopnia poniżej 65 roku życia

Dodatkowo AHA, wskazując na przyczyny miażdżycy tętnic wieńcowych, zwraca uwagę na wiek chorego oraz niespecyficzne zmiany odcinka ST w EKG spoczynkowym [55].

Badania naukowe dowodzą, że należy zwrócić szczególną uwagę na otyłość brzuszna, gdyż u osób cechujących się tym rodzajem otyłości częściej występuje hipercholesterolemia, hipertrójglicydemia, nadciśnienie tętnicze oraz zmiany w zapisie elektrokardiograficznym, wskazującym na występowanie ChNS [56, 57].

Dominacja współczulna AUN ma wpływ na pojawienie się niestabilności elektrycznej miokardium i może skutkować występowaniem groźnych arytmii [22, 58, 59]. Wiele dotychczas prowadzonych badań potwierdza istnienie związku pomiędzy obniżonymi wartościami parametrów zmienności rytmu serca a niedokrwieniem mięśnia sercowego u osób ostrej fazie zawału mięśnia sercowego (kiedy to dochodzi do wielu zmian w układzie autonomicznym), zwłaszcza w pierwszych dniach zawału [60, 61]. Uważa się jednak, że pomiary powinny być wykonane najwcześniej w 7 dobie od wystąpienia zawału [21]. Wraz z upływem czasu następuje stopniowy wzrost wartości parametrów zmienności rytmu serca. Proces ten rozpoczyna się po upływie ok. 2-3 tygodni od zawału i kończy się ok. 3 miesiąca, jednak parametry nie uzyskują wartości charakterystycznych dla osób zdrowych [62, 63].

Bigger i wsp. dowiedli, że w ostrej fazie zawału mięśnia sercowego dochodzi do obniżenia wartości wskaźników SDNN, RMSSD [64] i wartości współczynników mierzonych w dziedzinie częstotliwości [65]. Singh i wsp. [66] zauważyli obniżenie wartości wskaźnika LF/HF w ciągu 24 h u osób, które zmarły po upływie 30 dni od zawału. Shehab A. i wsp. twierdzą, że w wyniku ChNS dochodzi do obniżenia wartości wskaźnika SDNN [67]. Potwierdzają to inne badania, zgodnie z którymi obniżenie wartości SDNN poniżej 50 ms ma niekorzystne rokowanie [21, 68, 69, 65].

Zmiany w układzie autonomicznym w chorobie niedokrwiennej mięśnia sercowego mają związek z równoczesną obecnością czynników ryzyka miażdżycy. Wykazano istnienie związku pomiędzy obniżoną zmiennością a postępem lokalnych zmian miażdżycowych w naczyniach wieńcowych [70]. Kuzemczak i wsp. wskazali na obecność ujemnej korelacji pomiędzy wartościami wskaźników zmienności rytmu serca a nasileniem zmian miażdżycowych [16].

Europejskie Towarzystwo Kardiologiczne (European Society of Cardiology, ESC) zaleca, że do celów rokowniczych należy oceniać parametry zmienności serca z 24 godzinnych zapisów EKG [21]. Jednakże wiele badań dowodzi, że badania prowadzone w krótszych przedziałach czasowych są równie wiarygodne w przewidywaniu umieralności [48, 71, 72, 73].

2.2.2 Nadciśnienie a zmienność rytmu serca

Autonomiczny układ nerwowy odgrywa znaczącą rolę w regulacji ciśnienia krwi i rozwoju nadciśnienia tętniczego.

Niekontrolowane nadciśnienie tętnicze może prowadzić do poważnych chorób takich jak udar mózgu, choroba niedokrwienne z zawałem serca, niewydolność serca i nerek czy schorzenia siatkówki. Ryzyko pojawienia się tych powikłań, w wyniku wzrostu ciśnienia tętniczego, występuje stale, nawet jeśli wzrost ten mieści się w granicach wartości przyjmowanych jako prawidłowe [74, 75].

Tabela 3 Klasyfikacja nadciśnienia tętniczego według ESH/ESC [76]

Kategoria	Ciśnienie skurczowe (mmHg)	Ciśnienie rozkurczowe (mmHg)
Ciśnienie optymalne	< 120	< 80
Ciśnienie prawidłowe	120–129	80–84
Ciśnienie wysokie prawidłowe	130–139	85–89
Nadciśnienie stopień 1 (łagodne)	140–159	90–99
Nadciśnienie stopień 2 (umiarkowane)	160–179	100–109

Kategoria	Ciśnienie skurczowe (mmHg)	Ciśnienie rozkurczowe (mmHg)
Nadciśnienie stopień 3 (ciężkie)	≥ 180	≥ 110
Nadciśnienie izolowane skurczowe stopień 1	140 - 159	< 90
Nadciśnienie izolowane skurczowe stopień 2	≥ 160	< 90

Wg klasyfikacji amerykańskiej, wartości ciśnienia tętniczego zostają podzielone na 4 kategorie. Pojęcie „ciśnienia wysokiego” zostaje zastąpione pojęciem „stanu przednadciśnieniowego”. Określane są tylko 2 stopnie nadciśnienia tętniczego.

Tabela 4 Klasyfikacja nadciśnienia tętniczego według JNC 7 [77]

Kategoria	Ciśnienie skurczowe (mmHg)	Ciśnienie rozkurczowe (mmHg)
Ciśnienie prawidłowe	< 120	< 80
Stan przednadciśnieniowy	120–139	80–89
Nadciśnienie stopień 1	140–159	90–99
Nadciśnienie stopień 2	> 160	> 100

Obserwacje w warunkach spoczynku dowodzą, że u pacjentów z nadciśnieniem tętniczym dochodzi do obniżenia całkowitej zmienności rytmu serca oraz do przewagi aktywności układu sympatycznego [78]. Badania Framingham Heart Study [79] i Atherosclerosis Risk In Communities (ARIC) [80] wykazały związek obniżonej całkowitej HRV z zapadalnością na choroby układu sercowo-naczyniowego.

Wystąpienie patologii z obrębie układu krążenia, w tym nadciśnienia tętniczego, może być poprzedzane zaburzeniami równowagi współczulno-przywspółczulnej, i, co za tym idzie, odchyleniem wartości opisujących ją wskaźników HRV [81].

Jednakże badania HRV prowadzone wśród pacjentów z nadciśnieniem tętniczym nie dają jednoznacznych rezultatów. Pierwszymi badaczami analizującymi ten problem byli Mancina i wsp.. Nie wykazali oni istnienia istotnych różnic w HR oraz wartościach współczynnika SDNN wśród pacjentów z nadciśnieniem tętniczym w porównaniu z osobami zdrowymi [82].

Sprzeczne wyniki wykazał Radaelli i wsp. [83]. Wg nich osoby z nadciśnieniem cechowały się niższymi wartościami współczynnika SDNN, obliczanymi na podstawie wszystkich odstępów RR rytmu zatokowego w analizowanym zapisie, w porównaniu z osobami zdrowymi. Huikuri i wsp. [84] podobne zmiany zauważyli w grupie mężczyzn, chorujących na nadciśnienie tętnicze co najmniej 5 lat. Wykazali też, że czynnikiem najsilniej wpływającym na wartość SDNN jest skurczowe ciśnienie tętnicze. Zdecydowana większość badaczy wykazała zmniejszenie wahań, które zależą od aktywności układu przywspółczulnego — czyli składowej HF analizy częstotliwościowej HRV u pacjentów z nadciśnieniem tętniczym, zarówno granicznym [84, 85, 86, 87] i łagodnym [85] (przy czym pacjenci z łagodnym nadciśnieniem wykazywali silne zmiany w wartościach HF). Jednak inne badania tego nie potwierdzają [88, 89, 90]. Podobnie w stosunku do innych wskaźników uzyskano rozbieżne wyniki.

Tabela 5 Porównanie parametrów zmienności rytmu serca pomiędzy pacjentami z nadciśnieniem tętniczym a osobami zdrowymi [78]

Autorzy	Charakterystyka grupy	N	Wiek	Wyniki
Mancia i wsp. [82]	Nadciśnienie łagodne	26	40 ± 2	SDNN bez zmian
	Nadciśnienie ciężkie	41	45 ± 2	SDNN bez zmian
Radaelli i wsp. [83]	Nadciśnienie łagodne i umiarkowane	40	45 ± 2	SDNN ^a , LF ^a , HF ^a , LF/HF bez zmian
Huikuri i wsp. [84]	Nadciśnienie długo trwające (> 5 lat)	168	50 ± 6	SDNN ^{a,*} , LF ^{a,*}
Langewitz i wsp. [85]	Nadciśnienie graniczne	41	44 ± 8	HF ^{a,*}
	Nadciśnienie łagodne	34	44 ± 7	HF ^{a,*}
Piccirillo i wsp. [91]	Nadciśnienie łagodne	14	41 ± 9	LF ^{b,*} , LF/HF ^{b,*} , HF ^{a,*}
Guzzetti i wsp. [92]	Nadciśnienie niepowikłane	40	41 ± 3	LF ^{b,*} , LF/HF ^{b,*} , HF ^{a,*}
Yo i wsp. [88]	Nadciśnienie niepowikłane	31	21–71	HF, LF bez zmian
Takalo i wsp. [93]	Nadciśnienie graniczne	29	35–45	LF ^{b,*} , MF ^a

a – obniżenie w porównaniu z osobami zdrowymi, b – podwyższenie w porównaniu z osobami zdrowymi
* - wynik istotny statystycznie

Należy zwrócić uwagę na niewielkie liczebności badanych grup oraz występujące w nich niejednorodności, co mogło wpłynąć na różnice uzyskanych wyników.

2.2.3 Wpływ cukrzycy na zmienność rytmu serca

Zgodnie z definicją przyjętą pod koniec XX wieku, cukrzyca to grupa chorób metabolicznych, która charakteryzuje się hiperglikemią, powstającą w wyniku niewłaściwego wydzielania i/lub działania insuliny. Przewlekła hiperglikemia w cukrzycy jest związana z uszkodzeniem, zaburzeniem czynności i niewydolnością różnych narządów, szczególnie oczu, nerek, nerwów, serca i naczyń krwionośnych [94].

Rok rocznie liczba chorych na cukrzycę rośnie. Badania epidemiologiczne dowodzą systematycznego wzrostu przypadków tej choroby w ostatnim 10-leciu, czego skutkiem było określenie końca XX wieku mianem epidemii cukrzycowej. Zajmuje ona 6. miejsce pod względem częstości zgonów na świecie.

Choroba ta często nie daje żadnych objawów lub objawy są skąpe i dlatego błędnie traktowana jest jako choroba łagodna.

Cukrzyca, będąca czynnikiem ryzyka występowania choroby niedokrwiennej serca, powoduje dysfunkcję AUN. Wheeler i Ewing dowiedli, że powoduje ona obniżenie zmienności akcji serca [95, 96]. Silnie destabilizuje układ vegetatywny, co ma swoje odzwierciedlenie w obniżeniu zmienności rytmu serca (Liao i wsp. – ARIC Study) [97]. Dlatego też pacjenci z cukrzycą znacznie częściej zapadają na choroby sercowo-naczyniowe [98]. Mają również 2 – 4 razy większą szansę na wystąpienie choroby niedokrwiennej serca [99, 100]. U chorych na cukrzycę ryzyko zgonu z powodu choroby wieńcowej lub udaru mózgu jest 3-krotnie większe niż u osób bez cukrzycy [101]. Zauważono u nich szybszy rozwój zmian miażdżycowych w porównaniu z osobami bez cukrzycy, co powoduje, że mają większą szansę [102, 103] na wystąpienie nagłego epizodu wieńcowego oraz nagłej śmierci z przyczyn naczyniowo-sercowych.

Cukrzyca jest przyczyną występowania neuropatii, w tym, odnoszącej się do funkcjonowania serca, cukrzycowej neuropatii autonomicznej (Cardiac Autonomic Neuropathy, CAN). Objawami CAN mogą być m. in. zła tolerancja wysiłku, zaburzenia hormonalne czy gorsza adaptacja do podwyższonej temperatury. Jednym z wczesnych objawów klinicznych jest występowanie tachykardii podczas spoczynku. Nieprawidłowości nie dające żadnych objawów można stwierdzić już po roku od rozpoznania cukrzycy typu 2 i po 2 latach od

stwierdzenia cukrzycy typu 1. Choroba ta wiąże się istotnie z ryzykiem wystąpienia nagłego zgonu sercowego z powodu np. zawału mięśnia sercowego czy niedotlenienia. U pacjentów chorujących na cukrzycę typu 2 wykazano istnienie związku pomiędzy występowaniem choroby wieńcowej a CAN [104]. Zaburzenia w odczuwaniu niedotlenienia mięśnia sercowego, co skutkuje bezbólowym przebiegiem niedokrwienia, czyli tzw. „cichym” zawałem serca, również są efektem neuropatii. Cukrzycowa neuropatia autonomiczna ma związek z wydłużeniem skorygowanego odstępu QT [105]. Badania Khandokera i wsp. potwierdzają dysfunkcję kontroli autonomicznej tętna u pacjentów z CAN oraz wskazują, iż kompleksowa analiza zmienności rytmu serca może pomóc w identyfikacji pacjentów cierpiących na to schorzenie [106].

2.3 Wskaźniki zmienności rytmu serca

Wiele znanych i powszechnie stosowanych parametrów badających zmienność rytmu serca jest opartych na statystycznych miarach takich jak: średnia arytmetyczna, odchylenie standardowe, rozstęp, czy wreszcie rozstęp międzykwartyłowy. Część z nich wyznacza się bezpośrednio z odstępów RR, a niektóre opierają się na wartości HR, którą oblicza się stosując następujące przekształcenie:

$$HR = \frac{60000}{T_{RR}}$$

gdzie T_{RR} jest wyznaczane w milisekundach, a jako jednostkę HR przyjmuje się liczbę uderzeń na minutę (bpm) [31].

2.3.1 W dziedzinie czasu

Parametry określające zmienność rytmu serca mierzone w dziedzinie czasu można podzielić na trzy grupy:

- a) parametry zmienności krótkoterminowej, obrazujące zmiany w kolejnych przedziałach RR,
- b) parametry zmienności długoterminowej,
- c) parametry, które łączą w sobie zmienność krótko- i długoterminową.

W dziedzinie czasu analizowano 2 klasyczne parametry zmienności rytmu serca SDNN i RMSSD.

SDNN (standard deviation of NN) oblicza się jako odchylenie standardowe średnich wartości odstępów RR. Jest ono wyznaczone najczęściej na podstawie wszystkich 5-cio minutowych fragmentów EKG w badanym czasie.

$$SDNN = \sqrt{\frac{\sum_{i=1}^{N-1} T_{RR}^2(i)}{N-1}}$$

RMSSD to parametr wyznaczany jako pierwiastek II stopnia ze średniej z sumy kwadratów różnic kolejnych odstępów RR rytmu zatokowego w badanym czasie.

$$RMSSD = \sqrt{\frac{\sum_{i=1}^N [T_{RR}(i) - T_{RR}(i-1)]^2}{N-1}}$$

Obok klasycznych wskaźników zmienności rytmu serca poddano analizie jeszcze dodatkowe wskaźniki zmienności, zarówno krótko- jak i długoterminowej, które były do tej pory wykorzystywane m. in. w monitorowaniu częstości uderzeń serca płodu (Fetal Heart Rate, FHR).

Oscillation index

Należy do nich parametr zwany szerokością wstęgi oscylacyjnej (**Oscillation index**), który jest wrażliwy na wszelkie, zarówno krótko- jak i długoterminowe, zmienności. Oblicza się go jako rozstęp częstości rytmu serca i wyznacza mierząc sygnał w 60-cio sekundowych przedziałach czasowych z zachowaniem stałego sygnału (bez przyspieszeń i opóźnień).

$$OSC = \max(HR(i)) - \min(HR(i))$$

gdzie $i=1,2,\dots,N$

N - liczba uderzeń mierzona w 60-cio sekundowym oknie czasowym

Wskaźnik Yeh'a

Wskaźnik zmienności krótkoterminowej Yeh'a [107] (Yeh DI) jest definiowany jako odchylenie standardowe współczynników d_i , które wylicza się jako stosunek różnicy dwóch kolejnych interwałów $T_{RR}(i)$ i $T_{RR}(i+1)$ do ich sumy, otrzymywanych w 30 sekundowym oknie czasowym.

$$DI = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N-1} (d(i) - \bar{d})^2}$$

gdzie $d(i) = \frac{T_{RR}(i) - T_{RR}(i+1)}{T_{RR}(i) + T_{RR}(i+1)}$ oraz $\bar{d} = \frac{1}{N-1} \sum_{i=1}^{N-1} d(i)$

N – liczba uderzeń w ciągu 30s

Ocena zmienności długoterminowej wg Yeh'a jest określana jako iloraz odchylenia standardowego kolejnych interwałów $T_{RR}(i)$ względem średniej tych interwałów:

$$II = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (T_{RR}(i) - \bar{T})^2}}{\bar{T}}$$

gdzie $\bar{T} = \frac{1}{N} \sum_{i=1}^N T_{RR}(i)$

N – liczba uderzeń w ciągu 30s

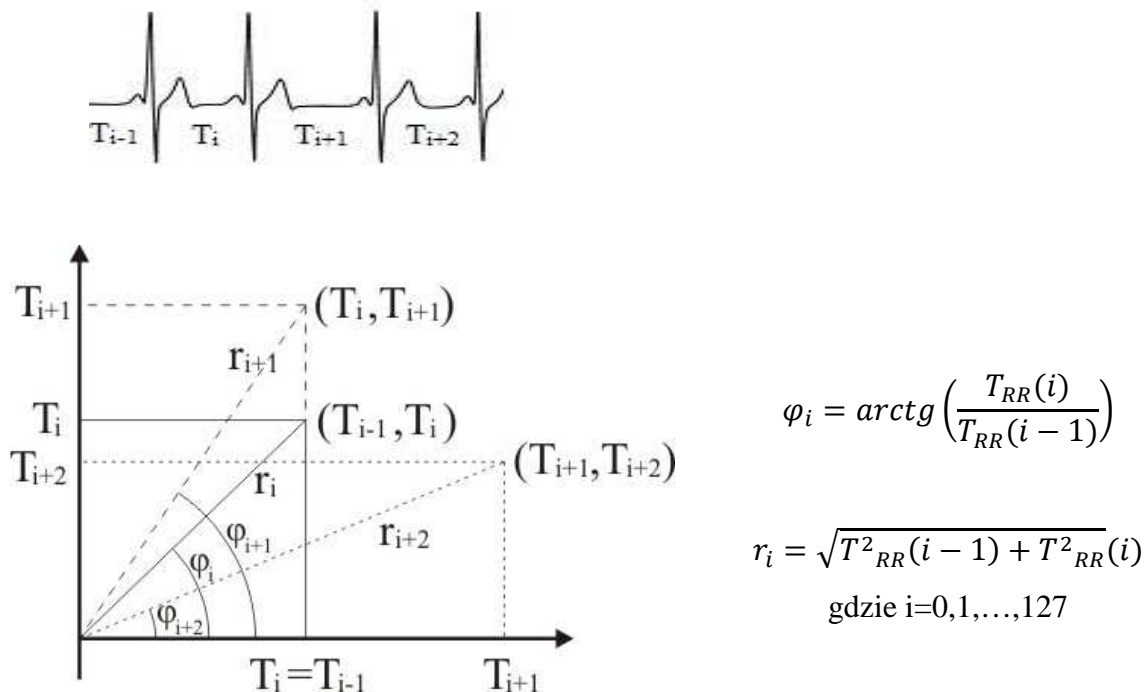
Wskaźnik de Haana

De Haan [108] zaproponował wykorzystanie biegunowego układu współrzędnych w celu graficznego przedstawienia rozkładu interwałów T_{RR} (ryc.1). W układzie współrzędnych zaznacza się pary kolejnych interwałów w postaci punktów. Określa się parametry takie, jak:

moduł (r_i) – odległość punktu od początku układu współrzędnych

φ (i) – kąt pod krzywą łączącą punkty z początkiem układu współrzędnych.

Obydwa wskaźniki de Haana odnoszą się do 128 kolejnych załamków R.



Wykres 2 Graficzny model rozkładu interwałów T_i w biegunowym układzie współrzędnych zaproponowany przez de Haana. Na podstawie J. Moczko „Advanced methods of heart rate signals processing and their usefulness in diagnosis suport. I. Mathematical heart rate descriptors and virtual instrumentation” [31]

Wskaźnik zmienności krótkoterminowej de Haana STI (Short Term Irregularity index) oblicza się jako rozstęp międzykwartyłowy argumentów φ_i . Zatem

$$STI=IQR(\varphi_i)$$

Zmienność długoterminowa (LTI) wg de Haana (Long Term Irregularity index), wyznaczana jest jako rozstęp międzykwartyłowy modułów r_i , czyli

$$LTI = IQR(r_i)$$

Wskaźnik Organa

Wskaźnik Organa [109] (Organ Band index) stosuje się wyznaczenia ilościowej zmienności długoterminowej rytmu serca. Jest on definiowany jako podwojone odchylenie standardowe częstości rytmu serca obserwowanych w ciągu kolejnych 30 s. Uważa się, że jest to przedział optymalny ze względu na wartość niesionej informacji o zmienności.

$$BAND = 2 \cdot \sqrt{\frac{1}{N-1} \sum_{i=1}^N (HR(i) - \overline{HR})^2}$$

gdzie \overline{HR} jest średnią arytmetyczną częstości rytmu serca w analizowanym oknie czasowym N – liczba uderzeń w ciągu kolejnych 30 s

Wskaźnik van Geijna

Analiza zmienności krótkoterminowej zaproponowana przez van Geijna [110] (van Geijn ID) oparta jest na rozstępie międzykwartyłowym ważonych różnic długości kolejnych interwałów T_{RR} , mierzonych w 30 sekundowym przedziale czasowym.

$$ID = IQR \{g_i(T_{RR}(i) - T_{RR}(i-1))\}$$

gdzie $g_i = \left(\frac{180}{\bar{T}_{RR}(i) - 320}\right)^{1.5}$ oraz $\bar{T}_{RR}(i) = \frac{T_{RR}(i) + T_{RR}(i-1)}{2}$

Dla $\bar{T}_{RR}(i) < 381ms$ g_i jest stałą i wynosi 5.

Wskaźnik Hueya

Model zaproponowany przez Hueya [111] jest oparty na chwilowej wartości tętna. Analizuje tendencje zmian jego wartości w czasie odpowiadającym stu okresom RR.

Wskaźnik zmienności krótkoterminowej wg Hueya (Huey STV) definiuje się jako sumę modułów różnic kolejnych wartości HR, pod warunkiem istnienia zmiany monotoniczności badanego sygnału.

$$Huey\ STV = \sum_{i=2}^{N-1} k \cdot |HR(i+1) - HR(i)|$$

$$\text{gdzie } k = \begin{cases} 1, & \text{gdzy } [HR(i-1) - HR(i)] \cdot [HR(i) - HR(i+1)] < 0 \\ 0, & \text{gdzy } [HR(i-1) - HR(i)] \cdot [HR(i) - HR(i+1)] \geq 0 \end{cases}$$

N=100

Zmienność długoterminową zaproponowaną przez Hueya określa się wówczas, jeżeli w przynajmniej trzech kolejnych uderzeniach serca nie odnotowano zmiany w monotoniczności sygnału HR (gdy przez przynajmniej trzy kolejne interwały nie zauważono zmiany znaku różnice sąsiednich wartości HR).

Jest ona obliczana jako suma bezwzględnych wartości różnic kolejnych HR, tzn.

$$Huey\ LTV = \sum_{i=2}^{N-1} l \cdot |HR(i+1) - HR(i)|$$

gdzie

$$l = \begin{cases} 1, & \text{gdzy } [HR(i-2) - HR(i-1)][HR(i-1) - HR(i)] \cdot [HR(i) - HR(i+1)] > 0 \\ 0, & \text{gdzy } [HR(i-2) - HR(i-1)][HR(i-1) - HR(i)] \cdot [HR(i) - HR(i+1)] \leq 0 \end{cases}$$

N=100

Wskaźnik Daltona

Dalton zaproponował dwa wskaźniki służące do ilościowej oceny zmian częstości uderzeń serca: MABB i SD [112].

Parametr zmienności krótkoterminowej (Dalton MABB) wyznacza się jako połowa średniej arytmetycznej bezwzględnych różnic dwóch następujących po sobie interwałów T_{RR} , ocenianych na podstawie 2 minutowej długości sygnału.

$$MABB = \frac{1}{2} \sum_{i=1}^N \frac{|T_{RR}(i) - T_{RR}(i+1)|}{N}$$

N – liczba interwałów w zapisie

Do wyznaczenia zmienności długoterminowej Dalton (Dalton SD) zaproponował zastosowanie współczynnika opartego na odchyleniu standardowym interwałów T_{RR} , zaobserwowanych w sygnale o długości 2 minut.

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (T_{RR}(i) - \bar{T}_{RR})^2}$$

gdzie $\bar{T}_{RR} = \frac{1}{N} \sum_{i=1}^N T_{RR}(i)$

N – liczba interwałów w sygnale o długości 2 minut

Wskaźnik Zugaiba

Model zaproponowany przez Zugaiba, do oceny zmienności krótkoterminowej wykorzystuje wartości interwałów czasowych T_i [113]. Wyznacza się pomocniczy parametr D_i definiowany jako stosunek bezwzględnej różnicy interwałów T_i do ich sumy, tzn.

$$D(i) = \frac{|T_{RR}(i+1) - T_{RR}(i)|}{T_{RR}(i+1) + T_{RR}(i)}$$

Wskaźnik zmienności krótkoterminowej (Zugaib STV) jest określany jako średnia arytmetyczna bezwzględnych różnic kolejnych wartości D_i i ich mediany w odniesieniu do 128 kolejnych okresów RR.

$$STV = \frac{1}{N-1} \sum_{i=1}^N |D(i) - Md[D(i)]|$$

gdzie $Md[D(i)]$ jest medianą 128 kolejnych wartości $D(i)$

Wskaźnik Zugaiba (Zugaib LTV) mówiący o zmienności długoterminowej jest definiowany jako przeciętne odchylenie interwałów T_{RR} względem mediany:

$$LTV = \frac{1}{N} \sum_{i=1}^N \frac{|T_{RR}(i) - Md[T_{RR}(i)]|}{Md[T_{RR}(i)]}$$

gdzie $Md[T_{RR}(i)]$ jest medianą wartości 128 interwałów $T_{RR}(i)$.

2.3.2 W dziedzinie częstotliwości

W celu wyznaczenia wskaźników zmienności rytmu zatokowego w dziedzinie częstotliwości zastosowano dwie metody: nieparametryczną, wykorzystującą algorytm Szybkiej Transformaty Fouriera (Fast Fourier Transformation, FFT) oraz parametryczną autoregresyjną.

Do wskaźników zmienności rytmu serca zmierzonych w dziedzinie częstotliwości, wyznaczonych przy użyciu FFT należą:

- ULF (ultra-low frequency range power) – zakres ultra niskich częstotliwości <DC; 0.003 Hz>
- VLF (very-low frequency range power) – zakres bardzo niskich częstotliwości (0.003 Hz; 0.04 Hz>

- LF (low frequency range power) – zakres niskich częstotliwości (0.04 Hz; 0.15 Hz)
- HF (high frequency range power) – zakres wysokich częstotliwości (0.15 Hz; μ Hz), gdzie górna granica tego przedziału różni się w zależności od wieku i wynosi:
 - $\mu=0.4$ Hz dla osób dorosłych
 - $\mu=0.8$ Hz dla noworodków
 - $\mu=1.1$ Hz dla płodów
- wskaźnik będący ilorazem LF/HF – uważany za wskaźnik modulacji współczulnej (wskaźnik równowagi współczulno-przywspółczulnej) [21, 114, 115].

Modelowanie autoregresyjne stanowiło alternatywną metodę analizy widma sygnału. Polega ono na zastosowaniu złożonych algorytmów obliczeniowych, w których kolejna próbka sygnału jest średnią arytmetyczną ważoną poprzednich. Dzięki tej metodzie można analizować sygnał złożony z niewielkiej liczby próbek. Zastosowanie jej umożliwia uzyskanie wysokiej rozdzielczości częstotliwościowej widma. Pozwala również na uzyskanie słabej zależności metody od stacjonarności sygnału oraz na uniezależnienie liczby zmiennych opisujących widmo od jego długości.

W pracy nie objęto analizą wskaźników otrzymanych za pomocą metody autoregresyjnej.

2.3.3 W połączonych dziedzinach czasu i częstotliwości

Do analizy zmienności rytmu serca w połączonych dziedzinach czasu i częstotliwości zastosowano technikę JTFA (Joint Time-Frequency Analysis) [37, 36]. Metoda ta po raz pierwszy została zaproponowana w połowie lat pięćdziesiątych przez mieszkającego w Anglii węgierskiego fizyka i laureata Nagrody Nobla Dennisa Gabora. Dalszymi pracami nad tą techniką zajęli się naukowcy z firmy National Instruments niezależnie z Xiaohong Zhang oraz Stéphane Mallat. Opracowane przez nich algorytmy, zakodowane w języku G, jako element pakietu Signal Processing Toolkit zostały dołączone do oprogramowania LabVIEW – środowiska, które daje możliwość tworzenia, testowania i uruchamiania aplikacji. Z dużą skutecznością może być ono również stosowane w przetwarzaniu sygnałów i zjawisk biomedycznych [116, 117, 118, 119].

W połączonych dziedzinach czasu i częstotliwości analizuje się również wskaźnik Allana oraz wskaźniki analizy falkowej $w_1 - w_5$.

Wskaźnik Allana

Wskaźnik Allana ($A(T)$) [31] jest oparty na transformacji falkowej Haar'a. Po raz pierwszy został zdefiniowany w 1996 roku. Określany jest jako stosunek wariancji Allana (zmiennosc kolejnych zliczeń $N_i(t)$ w z góry określonym przedziale czasowym T) do podwojonej wartości średniej.

$$A(T) = \frac{E\{[N_{i+1}(T) - N_i(T)]^2\}}{2E\{N_{i+1}(T)\}}$$

gdzie

$E\{\}$ – wartość oczekiwana (dla rozkładu normalnego jest ona równa wartości średniej arytmetycznej)

T – szerokość okna czasowego zliczającego załamki R

N_i – liczba zliczeń określająca liczbę załamków R znalezionych w analizowanym przedziale czasowym

Wskaźniki falkowe

Termin falka (ang. wavelet) został użyty po raz pierwszy przez Alfreda Haara'a w 1909 r. Technika ta opiera się na wyznaczeniu zestawu funkcji elementarnych skalowanych w czasie zwanych falkami. Wzrost parametru skalowania zmniejsza szerokość falki, co poprawia tym samym rozdzielczość czasową analizy z jednoczesnym pogorszeniem rozdzielczości częstotliwościowej [36, 31, 41].

$$\sigma_{wav}(m) = \left\{ E \left[|W_{m,n}(m)|^2 \right] \right\}^{\frac{1}{2}}$$

gdzie $E\{\}$ oznacza wartość oczekiwaną, a $W_{m,n}$ jest określany następująco:

$$W_{m,n}(m) = \frac{1}{\sqrt{m}} \sum_{i=1}^N T_{RR}(i) \psi\left(\frac{i-1}{m} - n\right)$$

gdzie

N – liczba interwałów T_{RR}

m – czynnik skalowania ($m=2^j$, gdzie j – indeks skali)

n – wartość translacji

Ψ – falkowa funkcja bazowa

2.3.4 Normalizacja wskaźników zmienności rytmu serca

Bezpośrednie porównywanie między sobą właściwości poszczególnych wskaźników zmienności rytmu serca, ze względu na definicje wskazujące sposób ich obliczania, jest niemożliwe. Wynika to przede wszystkim z różnicy w wymaganej szerokości okna czasowego, w którym rejestrowany jest sygnał. Aby uniknąć tego problemu można stosować dwa rodzaje parametrów:

- a) oryginalne – wyznaczone bezpośrednio z definicji;
- b) obliczane w znormalizowanym jednominutowym oknie czasowym [31].

Normalizacja oknem szerokości 1 minuty powoduje niewielkie zmiany w wartościach wskaźników: Yeha, de Haana, Organa, van Geijna, Hueya, Zugaiba, Daltona oraz szerokości wstęgi oscylacyjnej.

Tabela 6 Porównanie szerokości okna i związanej z nią liczby interwałów T_{RR} dla wybranych wskaźników HRV obliczonego dla średniej częstości rytmu serca wynoszącej 72 uderzenia/min [31]

Wskaźnik HRV	Definicja	Liczba odpowiadających interwałów T_{RR}	Liczba sekund
Yeh	Okno 30 s	36	30
de Haan	128 uderzeń	127	106
Organ	Okno 30 s	36	30
van Geijn	Okno 30 s	36	30
Huey	100 interwałów RR	100	83
Zugaib	128 interwałów RR	128	106
Dalton	Okno 120 s	100	120
OSC	Okno 60 s	72	60

2.3.5 Akwizycja danych (sygnałów biomedycznych)

Do rejestracji sygnałów uzyskanych podczas 5-minutowej rejestracji zapisu elektrokardiograficznego zastosowano system komputerowy KARDIOGRAF KARDIO PC (@MEDEA Gliwice). Następnie, po przesłaniu ich drogą elektroniczną i utworzeniu odpowiedniej bazy danych, poddano je analizie wykorzystując narzędzie wirtualne, utworzone w środowisku LabVIEW 7.1 Express firmy National Instruments, dzięki któremu możliwe było przeprowadzenie dokładnej, wielowymiarowej analizy danych, detekcji załamków R, oceny rytmu serca oraz estymacji widma przy użyciu deskryptorów mierzonych w dziedzinie czasu, częstotliwości i w połączonych dziedzinach [31].

Programowanie w środowisku programowym LabVIEW, jako, że jest to programowanie obiektowe, znacznie ogranicza możliwość popełnienia błędów składniowych oraz daje możliwość ich poprawy na poziomie wykonania programu, jednakże nie zmniejsza możliwości popełnienia przez programistę błędów logicznych. W skład programu napisanego w języku graficznym G w środowisku LabVIEW, zwanego narzędziem wirtualnym, wchodzi:

- panel czołowy – umożliwiający wprowadzanie do komputera danych wejściowych za pomocą urządzeń wejścia oraz na wyprowadzenie wyników; zawiera różnego rodzaju przełączniki, suwaki i pokręta,

- diagram – w którym zawarty jest kod źródłowy programu napisanego w języku G; jego elementami są: funkcje, narzędzia wirtualne niższego poziomu oraz struktury określające sposób i kolejność przetwarzania informacji,
- łączniki – służące do przekazywania danych pomiędzy programami i podprogramami
- ikony – będące symbolami utworzonego narzędzia wirtualnego; dzięki nim można odwoływać się do konkretnego narzędzia w zadaniach wyższego poziomu [36].

3. Cel pracy

Rozwój cywilizacji spowodował wzrost zachorowalności na choroby sercowo-naczyniowe. Szacuje się, że w Polsce zgony z powodu chorób krążenia stanowią 50% wszystkich zgonów.

Celem badania jest porównanie dwóch wybranych metod: regresji logistycznej, jako metody statystycznej, i drzew klasyfikacyjnych (jako metody DATA MINING) w celu określenia, która z nich lepiej opisuje osoby z przebyłym zawałem mięśnia sercowego, nadciśnieniem lub/i cukrzycą. W analizie uwzględnia się wskaźniki klasycznej analizy czasowej (SDNN i RMSSD), współczynniki zmierzone w dziedzinie czasu wyznaczone w oparciu o inne modele matematyczne, określające zmienność zarówno krótkoterminową (de Haana STI, Huey STV, Dalton MABB, Zugaib STV, Yeh DI, van Geijn) jak i długoterminową (de Haana LTI, Huey LTV, Dalton SD, Zugaib LTV, Yeh II, Organ BAND). Badane są również wskaźniki rytmu serca zmierzone w dziedzinie częstotliwości (LF/HF) oraz w połączonych dziedzinach czasu i częstotliwości (wskaźnik Allana). Dodatkowo zbadano wpływ takich czynników antropometrycznych jak płeć, wiek i BMI jako zmiennych wnikających w model.

Celem pracy jest znalezienie takich współczynników HRV oraz parametrów antropometrycznych, które mogą zostać zastosowane do charakterystyki osób z przebyłym zawałem mięśnia sercowego, nadciśnieniem lub/i cukrzycą.

Termin klasyfikacja, używany w dalszej części pracy, odnosi się do pojęcia matematycznego, powszechnie stosowanego do przyporządkowania obiektów do określonych klas i nie stanowi klasyfikacji w sensie medycznym, diagnostycznym.

4. Analizowane grupy

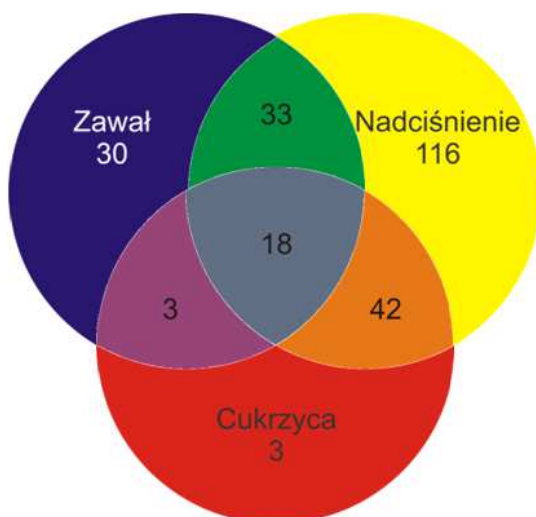
4.1 Grupa badana

Do badania zostali włączeni chorzy zakwalifikowani do planowego zabiegu koronarografii w stabilnym okresie choroby niedokrwiennej serca (tzw. dusznicy bolesnej stabilnej). Pacjenci byli hospitalizowani w Szpitalu Wojewódzkim w Poznaniu. Wszystkim chorym wykonano badanie przedmiotowe i podmiotowe oraz przeanalizowano dokumentację przez nich dostarczoną. Badanie było przeprowadzone w roku 2007, na jego przeprowadzenie uzyskano zgodę Komisji Bioetycznej Uniwersytetu Medycznego im. K. Marcinkowskiego w Poznaniu.

Do badania nie zakwalifikowano:

- pacjentów włączonych do badania ze wskazań innych niż dusznica bolesna;
- chorych ze świeżym zawałem mięśnia sercowego;
- pacjentów, u których w wyniku angioplastyki nie uzyskano zadowalającego przepływu w skali TIMI (Thrombolysis In Myocardial Infarction);
- chorych, u których z przyczyn technicznych nie uzyskano wystarczająco dobrej jakości zapisu elektrokardiograficznego umożliwiającego cyfrową obróbkę sygnału

Ostatecznie analizie poddano grupę badaną składającą się z 245 osób.



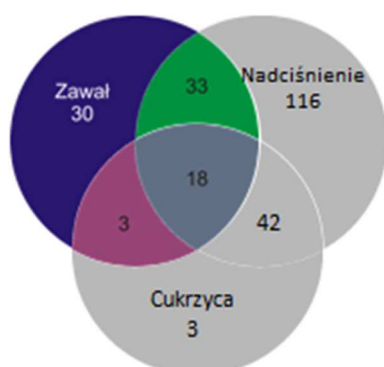
Wykres 3 Diagram Venna przedstawiający rozkład pacjentów w grupie badanej z uwzględnieniem jednostek chorobowych. Opracowanie własne.

W pracy zajęto się analizą czterech wybranych grup pacjentów:

- po przebytych zawale mięśnia sercowego (84 osoby) (4.1.1);
- po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie (33 osoby) (4.1.2);
- po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie i cukrzycę (18 osób) (4.1.3);
- chorującymi na nadciśnienie i cukrzycę (42 osoby) (4.1.4).

Grupę kontrolną stanowili pacjenci w stabilnym okresie choroby niedokrwiennej serca, zakwalifikowani do planowego zabiegu koronarografii, którzy nie przebyli zawału mięśnia sercowego oraz u których nie stwierdzono nadciśnienia i cukrzycy. Została ona opisana w punkcie 4.2.

4.1.1 Pacjenci po przebytych zawale mięśnia sercowego



Do badania włączono 84 chorych, którzy przebyli zawał mięśnia sercowego, 54 mężczyzn i 30 kobiet. Średnia wieku w grupie badanej wynosiła 59,0 lat (SD=9,7 lat). Najmłodsza osoba miała 35,0 lat, najstarsza 82,0 lata. Średnia wieku w grupie kobiet po przebytych zawale wynosiła 61,3 lat (SD=8,6 lat), a w grupie mężczyzn 58,0 lat (SD=10,0 lat). Najmłodsza kobieta miała 45,0 lat, najstarsza 76,0 lat. Najmłodszy mężczyzna po przebytych zawale mięśnia sercowego miał 35,0 lat, a najstarszy 82,0 lata. Ponad 50% wszystkich chorych miało co najmniej nadwagę. Wskaźnik masy ciała w tej grupie wynosił co najmniej 27 (Body Mass Index, BMI). Dokładną charakterystykę chorych po przebytych zawale mięśnia sercowego przedstawia poniższa tabela.

Tabela 7 Charakterystyka grupy z przebyłym zawałem mięśnia sercowego względem wieku i BMI, z uwzględnieniem podziału ze względu na płeć

	Cała grupa (N=84)			
	$\bar{x}\pm SD$	mediana (Q ₁ ; Q ₃)	min	max
wiek	59,0±9,7	57,0 (52,0; 66,5)	35,0	82,0
BMI	27,6±4,3	27,3 (24,6; 30,8)	18,7	38,4
Kobiety (N=30)				
	$\bar{x}\pm SD$	mediana (Q ₁ ; Q ₃)	min	max
wiek	61,3±8,6	59,0 (56,0; 70,0)	45,0	76,0
BMI	27,9±4,6	27,0 (23,7; 31,9)	20,4	37,0
Mężczyźni (N=54)				
wiek	58,0±10,0	56,0 (51,0; 66,0)	35,0	82,0
BMI	27,5±4,1	27,3 (24,7; 30,4)	18,7	38,3

Tabela 8 Charakterystyka grupy z przebyłym zawałem mięśnia sercowego względem palenia papierosów, podwyższonego poziomu cholesterolu i czynników ryzyka, z uwzględnieniem podziału ze względu na płeć

	Cała grupa (N=84)		Kobiety (N=30)		Mężczyźni (N=54)	
	Tak N (%)	Nie N (%)	Tak N (%)	Nie N (%)	Tak N (%)	Nie N (%)
Cukrzyca	21 (25,0)	63 (75,0)	11 (36,7)	19 (63,3)	10 (18,5)	44 (81,5)
Nadciśnienie	51 (60,7)	33 (39,3)	23 (76,7)	7 (23,3)	28 (51,9)	26 (48,1)

4.1.2 Pacjenci po przebyłym zawałe mięśnia sercowego i chorujący na nadciśnienie



Do badania włączono 33 chorych, którzy przebyli zawał mięśnia sercowego i mających nadciśnienie, 20 mężczyzn i 13 kobiet. Średnia wieku w grupie badanej wynosiła 60,5 lat (SD=10,4 lat). Najmłodsza osoba miała 35 lat, najstarsza 82,0 lata. Średnia wieku w grupie kobiet mających nadciśnienie i po przebyłym zawałe wynosiła 61,1 lat (SD=8,1 lat), a w grupie mężczyzn 60,1 lat (SD=11,9 lat). Najmłodsza kobieta miała 45,0 lat, najstarsza

74,0 lat. Najmłodszy mężczyzna z nadciśnieniem, po przebyłym zawałe mięśnia sercowego,

miał 35,0 lat, a najstarszy 82,0 lata. Ponad 50% wszystkich chorych miało co najmniej nadwagę (BMI ≥ 27). Dokładną charakterystykę chorych po przebytych zawale mięśnia sercowego chorujących na nadciśnienie przedstawia poniższa tabela.

Tabela 9 Charakterystyka grupy z przebytych zawalem mięśnia sercowego względem wieku i BMI, z uwzględnieniem podziału ze względu na płeć

	Cała grupa (N=33)			
	$\bar{x} \pm SD$	mediana (Q ₁ ; Q ₃)	min	max
wiek	60,5 \pm 10,4	59,0 (55,0; 68,0)	35,0	82,0
BMI	27,4 \pm 4,5	27,6 (23,7; 30,4)	20,3	38,4
Kobiety (N=13)				
wiek	61,1 \pm 8,1	59,0 (56,0; 69,0)	45,0	74,0
BMI	26,9 \pm 4,5	26,7 (23,2; 29,6)	20,4	33,7
Mężczyźni (N=20)				
wiek	60,1 \pm 11,9	59,0 (51,0; 67,5)	35,0	82,0
BMI	27,7 \pm 4,6	27,8 (24,3; 30,4)	20,3	38,4

4.1.3 Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie i cukrzycę



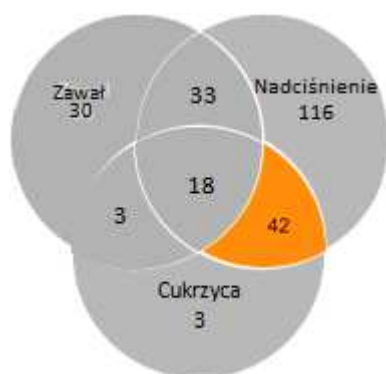
Do badania włączono 18 chorych, którzy przebyli zawal mięśnia sercowego i chorujących na nadciśnienie i cukrzycę, 8 mężczyzn i 10 kobiet. Średnia wieku w grupie badanej wynosiła 61,3 lat (SD=7,8 lat). Najmłodsza osoba miała 49,0 lat, najstarsza 74,0 lata. Średnia wieku w grupie kobiet mających nadciśnienie i cukrzycę oraz po przebytych zawale wynosiła 60,6 lat (SD=8,1 lat), a w grupie mężczyzn 62,3 lat (SD=6,5 lat). Najmłodsza kobieta miała 49,0 lat, najstarsza 74,0 lat. Najmłodszy mężczyzna po przebytych zawale mięśnia sercowego cierpiący na nadciśnienie i cukrzycę miał 54,0 lat, a najstarszy 72,0 lata. Ponad 50% wszystkich chorych miało co najmniej nadwagę (BMI ≥ 28).

Dokładną charakterystykę chorych po przebytych zawałach mięśnia sercowego chorujących na nadciśnienie i cukrzycę przedstawia poniższa tabela.

Tabela 10 Charakterystyka grupy z przebytych zawałami mięśnia sercowego, nadciśnieniem i cukrzycą względem wieku i BMI, z uwzględnieniem podziału ze względu na płeć

	Cała grupa (N=18)			
	$\bar{x}\pm SD$	mediana (Q ₁ ; Q ₃)	min	max
wiek	61,3±7,8	62,0 (55,0; 67,0)	49,0	74,0
BMI	29,8±4,3	28,1 (26,3; 33,7)	24,4	37,0
Kobiety (N=10)				
wiek	60,6±8,1	59,0 (52,0; 70,0)	49,0	74,0
BMI	30,8±4,5	31,6 (26,7; 34,6)	24,4	37,0
Mężczyźni (N=8)				
wiek	62,3±6,5	64,0 (55,5; 66,5)	54,0	72,0
BMI	28,4±3,8	27,4 (25,5; 31,2)	24,6	34,6

4.1.4 Pacjenci chorujący na nadciśnienie i cukrzycę



Do badania włączono 42 osoby, chorujące na nadciśnienie i cukrzycę, 23 mężczyzn i 19 kobiet. Średnia wieku w grupie badanej wynosiła 60,6 lat (SD=9,4 lat). Najmłodsza osoba miała 47,0 lat, najstarsza 81,0 lata. Średnia wieku w grupie kobiet mających nadciśnienie i cukrzycę wynosiła 60,4 lat (SD=9,7 lat), a w grupie mężczyzn 60,8 lat (SD=9,3 lat). Najmłodsza kobieta miała 47,0 lat, najstarsza 76,0 lat.

Najmłodszy mężczyzna chorujący na nadciśnienie i cukrzycę miał 47,0 lat, a najstarszy 81,0 lata. Ponad 50% wszystkich chorych miało co najmniej nadwagę (BMI $\geq 29,7$). Dokładną charakterystykę chorych na nadciśnienie i cukrzycę przedstawia poniższa tabela.

Tabela 11 Charakterystyka grupy z nadciśnieniem i cukrzycą względem wieku i BMI, z uwzględnieniem podziału ze względu na płeć

	Cała grupa (N=42)			
	$\bar{x}\pm SD$	mediana (Q ₁ ; Q ₃)	min	max
wiek	60,6±9,4	60,5 (54,0; 67,0)	47,0	81,0
BMI	30,0±4,4	29,7 (26,4; 33,5)	19,7	37,4
Kobiety (N=19)				
wiek	60,4±9,7	61,0 (50,0; 69,0)	47,0	76,0
BMI	29,7±4,2	29,2 (26,8; 33,6)	19,7	35,9
Mężczyźni (N=23)				
wiek	60,8±9,3	60,0 (54,0; 67,0)	47,0	81,0
BMI	30,2±4,6	29,8 (26,3; 33,5)	22,3	37,4

4.2 Grupa kontrolna

Grupę kontrolną stanowili pacjenci zakwalifikowani do planowego zabiegu koronarografii w stabilnym okresie choroby niedokrwiennej serca, u których jednak nie stwierdzono nadciśnienia i cukrzycy. Pacjenci ci nie przeżyli zawału mięśnia sercowego. Grupa złożona była z 62 osób, 25 kobiet i 37 mężczyzn. Średnia wieku w grupie kontrolnej wynosiła 56,1 lat (SD=10,2 lat). Najmłodsza osoba miała 35,0 lat, najstarsza 88,0 lata. Średnia wieku w grupie kobiet 58,0 lat (SD=11,9 lat), a w grupie mężczyzn 54,8 lat (SD=8,8 lat). Najmłodsza kobieta miała 41,0 lat, najstarsza 88,0 lat. Najmłodszy mężczyzna w tej grupie miał 35,0 lat, a najstarszy 81,0 lata. Ponad 50% wszystkich pacjentów miało co najmniej nadwagę (BMI $\geq 25,9$). Dokładną charakterystykę osób z grupy kontrolnej przedstawia poniższa tabela.

Tabela 12 Charakterystyka grupy kontrolnej względem wieku i BMI, z uwzględnieniem podziału ze względu na płeć

	Cała grupa (N=62)			
	$\bar{x}\pm SD$	mediana (Q ₁ ; Q ₃)	min	max
wiek	56,1±10,2	54,5 (51,0; 61,0)	35,0	88,0
BMI	25,8±2,8	25,9 (23,7; 27,8)	20,0	34,4
Kobiety (N=25)				
wiek	58,0±11,9	57,0 (51,0; 68,0)	41,0	88,0
BMI	25,0±3,2	25,2 (22,3; 27,6)	20,0	31,6
Mężczyźni (N=37)				
wiek	54,8±8,8	53,0 (51,0; 60,0)	35,0	81,0
BMI	26,3±2,5	26,2 (24,2; 27,8)	22,4	34,4

5. Metody

5.1 Metody jednowymiarowe

W badaniach medycznych często stosuje się analizy oparte na metodach jednowymiarowych, tzn. określających wpływ jednej zmiennej niezależnej (ilościowej lub jakościowej) na zmienną zależną. Dla zmiennych ilościowych jest to zazwyczaj test t-Studenta (lub test t-Studenta z poprawką Cochran-Coxa na brak jednorodności wariancji), który jest stosowany przy spełnieniu założenia normalności rozkładu badanej zmiennej lub test Manna-Whitneya (stosowany w przypadku braku spełnienia tego założenia). Do najczęściej stosowanych testów służących do analizy danych jakościowych należy test niezależności chi-kwadrat wraz z poprawkami: Yatesa, testem dokładnym Fishera lub jego rozszerzeniem na tabele większe niż 2x2 – testem Fishera-Freemana-Haltona. Warunkiem zastosowania tych testów jest przedstawienie danych w postaci tzw. tabeli kontyngencji, tj. tabeli o następującej budowie:

Tabela 13 Przykładowa tabela kontyngencji

Liczebności obserwowane		choroba tak	choroba nie	suma
Narażenie na czynnik	tak	a	b	a + b
	nie	c	d	c + d
	suma	a + c	b + d	n = a + b + c + d

Według Cochran [120] test chi-kwadrat można stosować wtedy, gdy spełnione są następujące warunki:

- żadna z liczebności oczekiwanych nie może być < 1 ;
- nie więcej niż 20% liczebności oczekiwanych może być < 5 ,

gdzie liczebności oczekiwane wyznaczone są z następującego wzoru:

$$t_{ij} = (\text{suma } i\text{-tego wiersza} * \text{suma } j\text{-tej kolumny}) / n$$

t_{ij} – liczebność oczekiwana odpowiadająca liczebności obserwowanej n_{ij} , tzn. leżącej na przecięciu i -tego wiersza i j -tej kolumny

Test dokładny Fishera jest dedykowany dla tabel 2x2 (tzn. mających 2 wiersze i 2 kolumny) o małych liczebnościach oczekiwanych [121, 122], co wynika z przede wszystkim ze sposobu obliczania testu dokładnego Fishera, opierającego się na silni:

$$P = \frac{\binom{a+c}{a} \binom{b+d}{d}}{\binom{n}{a+b}} = \frac{\frac{(a+c)!}{a!c!} * \frac{(b+d)!}{d!b!}}{\frac{n!}{(a+b)!(n-a-b)!}}$$

W przypadku potwierdzenia istnienia zależności, często wyznacza się również iloraz szans (Odds Ratio, OR) wraz z podaniem przedziałów ufności.

W badaniu zastosowano test Manna-Whitneya (w celu wykrycia istnienia pomiędzy badanymi grupami ewentualnych różnic w wartościach poszczególnych parametrów) oraz test niezależności chi-kwadrat (w celu zbadania istnienia zależności pomiędzy analizowanymi zmiennymi).

5.2 Metody wielowymiarowe

Pod pojęciem analizy wielowymiarowej rozumie się grupę metod statystycznych pozwalających na dokonanie jednoczesnej analizy pomiarów co najmniej trzech zmiennych. Przykładem analizy wielowymiarowej jest badanie jednoczesnego wpływu kilku zmiennych niezależnych (lub wyjaśniających) na zmienną zależną (lub wyjaśnianą). Metodą, często stosowaną do tego typu analiz jest regresja logistyczna.

5.2.1 Regresja logistyczna

Pierwsze prace dotyczące zastosowania funkcji logistycznej powstały już pod koniec XIX w., kiedy to m. in. P.F. Verhulst [123] wykorzystał ją do opisu właściwości demograficznych. Podstawy metody regresji logistycznej zostały opracowane w 1944r. przez Berksona [124], natomiast jej pełny model po raz pierwszy zastosował w 1972 r. D.J. Finney.

Regresja logistyczna (logistic regression, LR) jest to model matematyczny, służący do opisanego wpływu jednej lub kilku zmiennych niezależnych $x_1, x_2, x_3, \dots, x_n$ na dychotomiczną zmienną zależną y . Pozwala on na włączenie do modelu zarówno zmiennych ilościowych jak i jakościowych. W regresji logistycznej zwykle zmienna zależna jest dychotomiczna, tzn. posiada dwa stany: występowanie lub brak danej cechy. Można również zastosować model regresji dla zmiennej zależnej o więcej niż dwóch kategoriach (mówi się wówczas o multinominalnej regresji logistycznej) lub dla zmiennej zależnej porządkowej (porządkowa regresja logistyczna).

Regresja logistyczna znajduje zastosowanie w wielu dziedzinach: demografii, epidemiologii (rozprzestrzenianie się epidemii), psychologii (ocena postępów w nauce), marketingu oraz medycynie, ponieważ za jej pomocą można określać ryzyko znalezienia się jednostki badania w pewnym stanie oraz znaleźć statystycznie istotne czynniki mające na to ryzyko wpływ. Popularność metody wynika z tego iż w prosty sposób wiąże parametry modelu z ilorazem szans, który jest łatwy w interpretacji. Dodatkowym atutem tego modelu jest możliwość interpretacji jego parametrów.

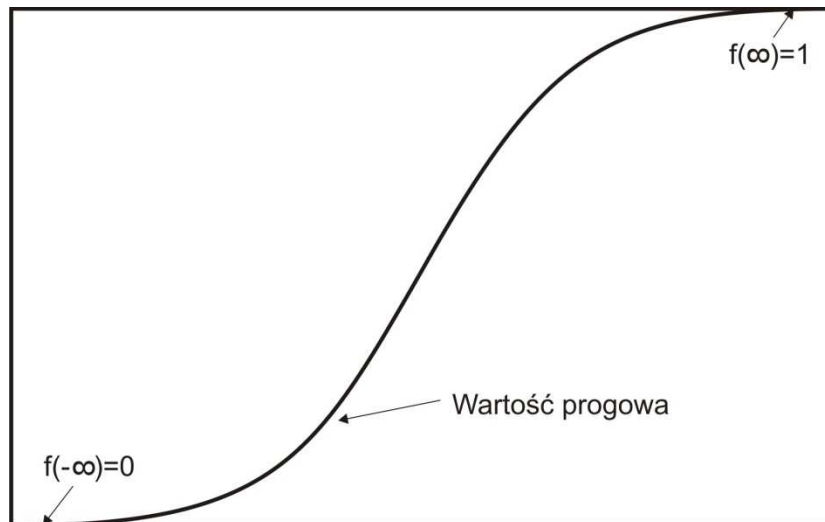
a) postać funkcji logistycznej

Funkcja logistyczna $f(z)$ określana jest wzorem

$$f(z) = \frac{e^z}{1 + e^z}$$

gdzie e - liczba Eulera, $e \approx 2,71828$.

Przyjmuje ona wartości należące do przedziału $\langle 0,1 \rangle$, przy czym 0 i 1 są wartościami brzegowymi osiąganymi odpowiednio w $-\infty$ i w $+\infty$.



Wykres 4 Postać funkcji logistycznej, opracowanie własne na podstawie A. Stanisław [125]

Wykres funkcji logistycznej ma kształt wydłużonej litery S. Dokonując jego analizy można zauważyć, że początkowo wartości funkcji, dla wzrastających argumentów x, rosną bardzo powoli, aż do osiągnięcia wartości progowej. Po jej przekroczeniu wartości funkcji zaczynają gwałtownie rosnąć, by ponownie ustabilizować się około wartości 1 (do której zbliża się asymptotycznie w $+\infty$). Funkcja jest szczególnie przydatna w analizie danych kategoryalnych, gdyż może opisywać prawdopodobieństwo wystąpienia bądź nie pewnego zdarzenia (prawdopodobieństwo przyjmuje wartości od 0 do 1).

b) model logistyczny

Model regresji logistycznej dla zmiennej dychotomicznej definiowany jest równaniem:

$$P(Y = 1 | x_1, x_2, \dots, x_n) = \frac{e^{\left(a_0 + \sum_{i=1}^n a_i x_i\right)}}{1 + e^{\left(a_0 + \sum_{i=1}^n a_i x_i\right)}}$$

gdzie:

Y – to zmienna zależna dychotomiczna, przyjmująca wartości:

$$Y = \begin{cases} 1, & \text{gdy zdarzenie wystąpiło} \\ 0, & \text{gdy zdarzenie nie wystąpiło} \end{cases}$$

$x_1, x_2, x_3, \dots, x_n$ – zmienne niezależne (predyktory);

$a_1, a_2, a_3, \dots, a_n$ – współczynniki regresji odpowiadające zmiennym $x_1, x_2, x_3, \dots, x_n$

$P(Y = 1 \mid x_1, x_2, x_3, \dots, x_n)$ oznacza warunkowe prawdopodobieństwo osiągnięcia przez zmienną zależną wartości 1, przy konkretnych wartościach zmiennych $x_1, x_2, x_3, \dots, x_n$.

Za pomocą równania regresji logistycznej, podobnie jak równania regresji liniowej, oblicza się wartość oczekiwaną zmiennej zależnej [126], przy czym, w regresji logistycznej dla dwukategoryjnej zmiennej zależnej, wartość oczekiwana zmiennej Y została zastąpiona warunkowym prawdopodobieństwem przyjęcia przez zmienną zależną wartości 1, dla zmiennych niezależnych $x_1, x_2, x_3, \dots, x_n$.

c) metody estymowania parametrów

Aby dopasować jak najlepszy model, bazując na wartościach pewnej grupy zmiennych, należy oszacować wartości współczynników regresji $a_0, a_1, a_2, \dots, a_n$. Należy zatem pamiętać, że liczebność badanej grupy (n) powinna spełniać warunek:

$$n > 10(k + 1),$$

gdzie k – oznacza liczbę parametrów włączonych do modelu [125].

Współczynniki regresji logistycznej estymowane są metodą najwyższej wiarygodności (maximum likelihood). Nie można w tym przypadku zastosować metody najmniejszych kwadratów, gdyż wymaga ona założenia o stałości wariancji, które nie jest spełnione w przypadku zmiennej dychotomicznej. Podstawy metody najwyższej wiarygodności zostały opracowane przez Fishera w 1929 r. Algorytm obliczeniowy tej metody opiera się na wielokrotnym estymowaniu wszystkich współczynników regresji, tak, by maksymalizować funkcję wiarygodności, określoną wzorem:

$$L = \prod_{i=1}^n P(y_i \mid a_1, a_2, \dots, a_k)$$

gdzie $P(y_i \mid a_1, a_2, \dots, a_k)$ oznacza prawdopodobieństwo pojawienia się wartości y_i przy danym modelu regresji z parametrami a_1, a_2, \dots, a_k [127, 125, 128], czyli maksymalizuje iloczyn prawdopodobieństw pojawienia się poszczególnych obserwacji z próby przy danych parametrach modelu. Często funkcję wiarygodności zastępuje się jej logarytmem, z racji tego,

że łatwiej, z przyczyn obliczeniowych, znaleźć maksimum funkcji $\log(L)$, niż samej funkcji L [129, 125, 128].

Ideą metody największej wiarygodności jest to, żeby przyjąć jako oceny szacowanych parametrów te wartości, których wiarygodność jest największa. Tak wyznaczone oceny gwarantują największe prawdopodobieństwo otrzymania zaobserwowanych wartości zmiennej zależnej.

d) testowanie hipotez w modelu regresji logistycznej

Stosując metodę najwyższej wiarygodności otrzymujemy, obok wartości współczynnika największej wiarygodności (L), listę zmiennych w modelu wraz z wyznaczonymi współczynnikami regresji oraz wartościami błędów standardowych.

Testowanie hipotez w modelu regresji logistycznej ma na celu odpowiedź na pytanie, czy model, który zawiera pewną zmienną dostarcza istotnie więcej informacji o zmiennej objaśnianej od modelu, który tej zmiennej nie zawiera.

Są dwa podejścia do testowania istotności współczynników regresji:

- obliczanie ilorazu wiarygodności (likelihood ratio) dla całego modelu;
- obliczenie wartości statystyki Walda dla każdej składowej z osobna, dzięki czemu można ocenić, które z wprowadzonych zmiennych niezależnych i w jakim stopniu istotnie wpływają na zmienną zależną.

Iloraz wiarygodności uwzględnia istotność całego modelu, a nie tylko pojedynczych parametrów. Daje odpowiedź na pytanie, czy model zawierający zmienną (bądź zmienne) niezależną daje lepsze przewidywanie wyników w porównaniu z modelem, który tej zmiennej nie zawiera. Zwykle modelem odniesienia jest model zerowy, tzn. zawierający tylko wyraz wolny.

Testowanie istotności zmiennej niezależnej opiera się na statystyce dewiancji (deviance, D), określonej następująco [127]:

$$D = -2 \log \left[\frac{\text{wartość funkcji wiarygodności estymowanego modelu}}{\text{wartość funkcji wiarygodności modelu pełnego}} \right]$$

przy czym pod pojęciem modelu pełnego rozumie się model o liczbie parametrów równej liczbie obserwacji.

W modelu regresji logistycznej wartość funkcji wiarygodności modelu pełnego jest równa 1. Zatem

$$D = -2 \log [\text{wartość funkcji wiarygodności estymowanego modelu}].$$

Ocenę istotności zmiennej objaśnianej przeprowadza się na podstawie statystyki G ilorazu wiarygodności. Mierzy ona zmianę wartości dewiancji w wyniku dodania do modelu zmiennej, której istotność jest badana:

$$G = D(\text{model bez badanej zmiennej}) - D(\text{model zawierający badaną zmienną}) \\ = -2 \log \left[\frac{L(\text{model bez badanej zmiennej})}{L(\text{model zawierający badaną zmienną})} \right]$$

Dla dostatecznie dużej liczby obserwacji (n) statystyka G ma rozkład zbliżony do rozkładu chi-kwadrat z liczbą stopni swobody równą różnicy między liczbą parametrów w obu analizowanych modelach (dla powyższego przykładu liczba stopni swobody wynosi 1).

Zwykle porównuje się modele różniące się od siebie jedną zmienną niezależną w celu sprawdzenia, czy dodana zmienna istotnie zwiększa trafność przewidywania modelu [129]. Przy budowie modelu stosowane są dwa podejścia: selekcja postępująca (budowę modelu zaczyna się od modelu zerowego, a następnie włącza kolejno zmienne niezależne) lub eliminacja wsteczna (zaczyna się od modelu pełnego i eliminuje kolejno zmienne, które są nieistotne).

Współczynnik Z-Walda stosowany jest do testowania hipotez zerowych dla współczynników regresji logistycznej dla każdej zmiennej w modelu, tzn. hipotezę $H_0: a_i = 0$, dla zmiennej x_i . Hipoteza ta mówi o braku wpływu predykatora na zmienną zależną. Współczynniki regresji są estymowane z pewną dokładnością (zwaną błędem standardowym) i oznaczaną SE_{a_i} .

Przyjmuje się, że badana zmienna x_i jest istotna statystycznie, jeżeli wartość jej współczynnika (a_i) jest różna od 0 oraz jest on wyznaczony z odpowiednio małym błędem

SE_{a_i} , zgodnie ze statystyką Walda, opierającą się na wyznaczonych współczynnikach regresji oraz ich błędach standardowych [127], tzn.

$$Z = \frac{a_i}{SE_{a_i}}$$

Rozkład współczynnika Walda jest w przybliżeniu zgodny z rozkładem normalnym dla dużych prób, natomiast rozkład współczynnika Walda podniesionego do 2 potęgi jest zgodny z rozkładem chi-kwadrat z jednym stopniem swobody. W większości pakietów statystycznych współczynnik Walda (Z) przedstawiany jest w postaci podniesionej do kwadratu, stąd też często mówi się o współczynniku chi-kwadrat Walda.

e) iloraz szans

W modelu regresji logistycznej, obok współczynników regresji i ich statystycznej istotności, ważną rolę odgrywa iloraz szans. Parametr ten jest często stosowany w badaniach klinicznych lub epidemiologicznych.

Szansa jest funkcją prawdopodobieństwa. Zamiast obliczania stosunku liczby sukcesów do liczby wszystkich prób (czyli klasycznego prawdopodobieństwa), wylicza się stosunek prawdopodobieństwa, że dane zjawisko wystąpi (sukces), do prawdopodobieństwa, że dane zjawisko nie wystąpi (porażka). Niech $S(A)$ oznacza szansę, a $p(A)$ prawdopodobieństwo sukcesu. Wówczas szansę można opisać wzorem:

$$S(A) = \frac{p(A)}{1 - p(A)}$$

Prawdopodobieństwo zdarzenia p przyjmuje wartości $\langle 0,1 \rangle$, więc szansa przyjmuje wartości należące do przedziału $(0, +\infty)$.

Iloraz szans odnosi się do sytuacji, gdy występowanie danego zjawiska badane jest w dwóch niezależnych i porównywalnych grupach, np. grupa A – grupa narażona na czynnik, grupa B – grupa nienarażona na czynnik. Definiowany jest jako stosunek szansy wystąpienia tego zjawiska w grupie A do szansy jego wystąpienia w grupie B, tzn.

$$OR_{AxB} = \frac{S(A)}{S(B)} = \frac{p(A)}{1 - p(A)} \div \frac{p(B)}{1 - p(B)}$$

gdzie $OR_{A \times B}$ oznacza iloraz szans dla grup A i B.

W przypadku danych dychotomicznych, które można przedstawić w postaci następującej tabeli

	narażenie na czynnik		
	tak	nie	suma
choroba tak	a	b	a + b
choroba nie	c	d	c + d
suma	a + c	b + d	n = a + b + c + d

iloraz szans oblicza się jako stosunek szansy wystąpienia danego zdarzenia w grupie narażonych na czynnik do szansy wystąpienia tego zdarzenia w grupie osób nienarażonych na czynnik.

Szansa wystąpienia zdarzenia w grupie osób narażonych wynosi:

$$S(\text{narażona}) = \frac{\frac{a}{a+c}}{\frac{c}{a+c}} = \frac{a}{c}$$

Szansa wystąpienia zdarzenia w grupie osób nienarażonych określana jest wzorem:

$$S(\text{nienarażona}) = \frac{\frac{b}{b+d}}{\frac{d}{b+d}} = \frac{b}{d}$$

Dla wyznaczonych szans, iloraz szans oblicza się następująco:

$$OR_{A \times B} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \cdot d}{b \cdot c}$$

Zastosowanie w przypadku modelu regresji logistycznej transformacji logitowej, tj. przekształcenia

$$\text{logit } P = \ln\left(\frac{P}{1-P}\right)$$

daje następującą równość:

$$\text{logit } P = a_0 + \sum_{i=1}^n a_i x_i$$

Zgodnie z definicją logarytm szansy wystąpienia zdarzenia A wynosi:

$$\ln(S(A)) = \ln\frac{P(Y=1)}{1-P(Y=1)} = \text{logit } P(Y=1) = a_0 + \sum_{i=1}^n a_i x_i$$

Stąd szansa wynosi:

$$S(A) = e^{a_0 + \sum_{i=1}^n a_i x_i}$$

Zatem, aby obliczyć iloraz szans dla dwóch porównywalnych grup należy zastosować wzór:

$$OR_{A \times B} = e^{\sum_{i=1}^n a_i (x_{A_i} - x_{B_i})}$$

gdzie

x_{A_i} - określa wartość zmiennej x_i w grupie A

x_{B_i} - określa wartość zmiennej x_i w grupie B

Iloraz szans interpretujemy następująco:

- jeżeli $OR_{A \times B} > 1$, to w szansa wystąpienia danego zdarzenia w grupie A jest większa niż w grupie B;
- jeżeli $OR_{A \times B} = 1$, to w obu grupach szansa na wystąpienie danego zdarzenia jest taka sama;
- jeżeli $OR_{A \times B} < 1$, to szansa na wystąpienie danego zdarzenia w grupie A jest mniejsza niż w grupie B.

f) dobroć dopasowania modelu regresji logistycznej

Testowanie hipotez w modelu regresji logistycznej, oparte zarówno na analizie ilorazu wiarygodności jak i współczynnika Walda, ma na celu udzielenie odpowiedzi na pytanie, czy model zawierający określone predyktory powoduje trafniejsze przewidywanie wartości zmiennej zależnej w porównaniu z modelem zawierającym mniej zmiennych niezależnych (lub inne zmienne niezależne – czasami tą samą zmienną zależną można opisać za pomocą kilku różnych modeli, zbudowanych z różnych zmiennych). Nie są to jednak miary mówiące o dopasowaniu modelu do danych obserwowanych. Pewną pomocą w odpowiedzi na pytanie dotyczące dopasowania modelu może być informacja o tym, jaki procent obserwacji z próby zostanie poprawnie zaklasyfikowany na podstawie modelu. Należy pamiętać jednak o tym, że nie można jednoznacznie stwierdzić, jaki procent poprawnie zaklasyfikowanych przypadków oznacza dobre dopasowanie modelu [130].

Procedura zaproponowana przez Hosmera i Lemeshowa opiera się na teście chi-kwadrat. Najpierw dokonuje sortowania rosnąco obserwacji z próby względem, wyestymowanego na podstawie modelu, prawdopodobieństwa przyjęcia przez zmienną zależną wartości 1. Następnie przypadki są dzielone na g grup. Dla tabel kontyngencji o wymiarach $g \times 2$, tzn. g wierszach i 2 kolumnach (w których znajdują się częstości rzeczywiste i przewidywane na podstawie modelu) oblicza się wartość współczynnika chi-kwadrat. Z racji tego, że test dobroci dopasowania Hosmera i Lemeshowa porównuje wartości obserwowane z wartościami oczekiwanymi na podstawie modelu, oczekuje się uzyskania braku istotności współczynnika chi-kwadrat.

Istnieją również inne statystyki opisujące, jak bazując na zmiennych niezależnych, można dobrze określić zmienną zależną. W analizie regresji prostej i wielorakiej informację taką niesie współczynnik determinacji (R^2), który mierzy, jaka część ogólnej zmienności zmiennej zależnej jest wyjaśniana regresją liniową [125]. W przypadku regresji logistycznej odpowiednikiem współczynnika determinacji jest m. in., zaproponowany w 1974 r. przez McFaddena, współczynnik pseudo R^2 określony wzorem:

$$R_{McFadden}^2 = 1 - \frac{\ln L_p}{\ln L_0}$$

gdzie $\ln L_p$ jest logarytmem funkcji wiarygodności dla modelu pełnego;

$\ln L_0$ jest logarytmem funkcji wiarygodności dla modelu zawierającego tylko wyraz wolny.

Im większa jest wartość współczynnika pseudo R^2 , tym lepsze jest przewidywanie zmiennej za pomocą naszego modelu. Wartość pseudo $R^2 = 0$ oznacza, że wszystkie parametry uwzględnione w modelu, poza wyrazem wolnym, są nieistotne.

W pracy obliczono współczynnik Nagelkerke określony wzorem:

$$\bar{R}^2 = \frac{R_{McFadden}^2}{\max(R_{McFadden}^2)} = \frac{1 - e^{-(2/n)(\ln L_p - \ln L_0)}}{1 - e^{(2/n)\ln L_0}}$$

Jest on modyfikacją współczynnika pseudo R^2 , tak by jako wartość maksymalną przyjmował wartość jeden.

Istnieją również inne współczynniki (np. Coxa-Snella), które zostały opisane przez A. Stanisza [125].

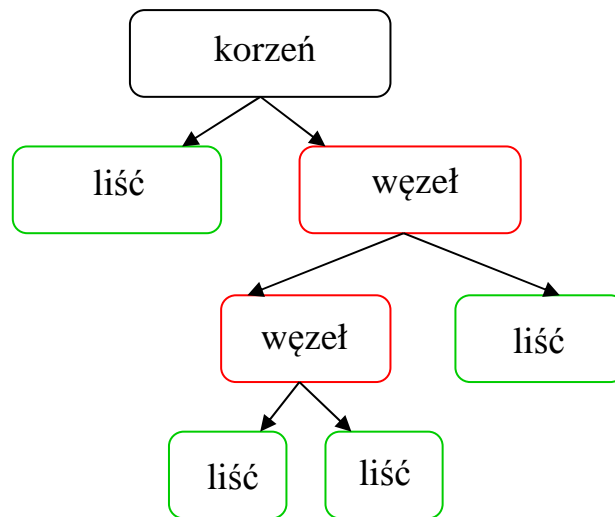
5.2.2 Drzewa klasyfikacyjne

Drzewa klasyfikacyjne (często zwane również decyzyjnymi) są wykorzystywane w wielu dziedzinach nauki. Znajdują swoje zastosowanie zarówno w nauczaniu maszynowym jak i w teorii prawdopodobieństwa i statystycznym rozpoznawaniu obrazów. Wykorzystuje się je w dziedzinach stosowanych, np. w medycynie, botanice czy psychologii. Analiza drzew klasyfikacyjnych jest jedną z podstawowych technik stosowaną w tzw. Zgłębianiu Danych (DATA MINING). Wykorzystuje się je do wyznaczania przynależności przypadków lub obiektów do klas jakościowej zmiennej zależnej na podstawie pomiarów jednej lub więcej zmiennych objaśniających (predyktorów). Zastosowane do ich budowy algorytmy uczenia się, cieszą się dużą popularnością. Wynika to z tego, że za pomocą drzewa może być przedstawiona dowolna funkcja odwzorowująca wartości wszystkich określonych na dziedzinie atrybutów na zbiór kategorii. Struktura drzew pozwala na budowę ogólnych reguł, umożliwia także bardzo efektywną implementację procesu klasyfikacji. Elastyczność tej analizy powoduje to, że jest ona coraz chętniej stosowana. Jest nieoceniona w przypadku, gdy zawodzą tradycyjne metody. Należy jednak pamiętać o tym, że jeśli założenia metod tradycyjnych są spełnione, zaleca się ich wykorzystanie, zamiast, mniej restrykcyjnych, drzew klasyfikacyjnych. Kolejnym powodem powszechności drzew jest to, iż dają się łatwo przedstawić graficznie, przez co są łatwiejsze w interpretacji niż wyniki liczbowe. Drzewa klasyfikacyjne czasami mają złożoną budowę. Jednak możliwość graficznego przedstawienia ich struktury ułatwia interpretację nawet skomplikowanych rozwiązań [131, 132].

Celem analizy przeprowadzanej w oparciu o drzewa jest przewidywanie wartości jakościowej zmiennej zależnej na podstawie zmiennych niezależnych lub wyjaśnianie zakodowanej w niej odpowiedzi .

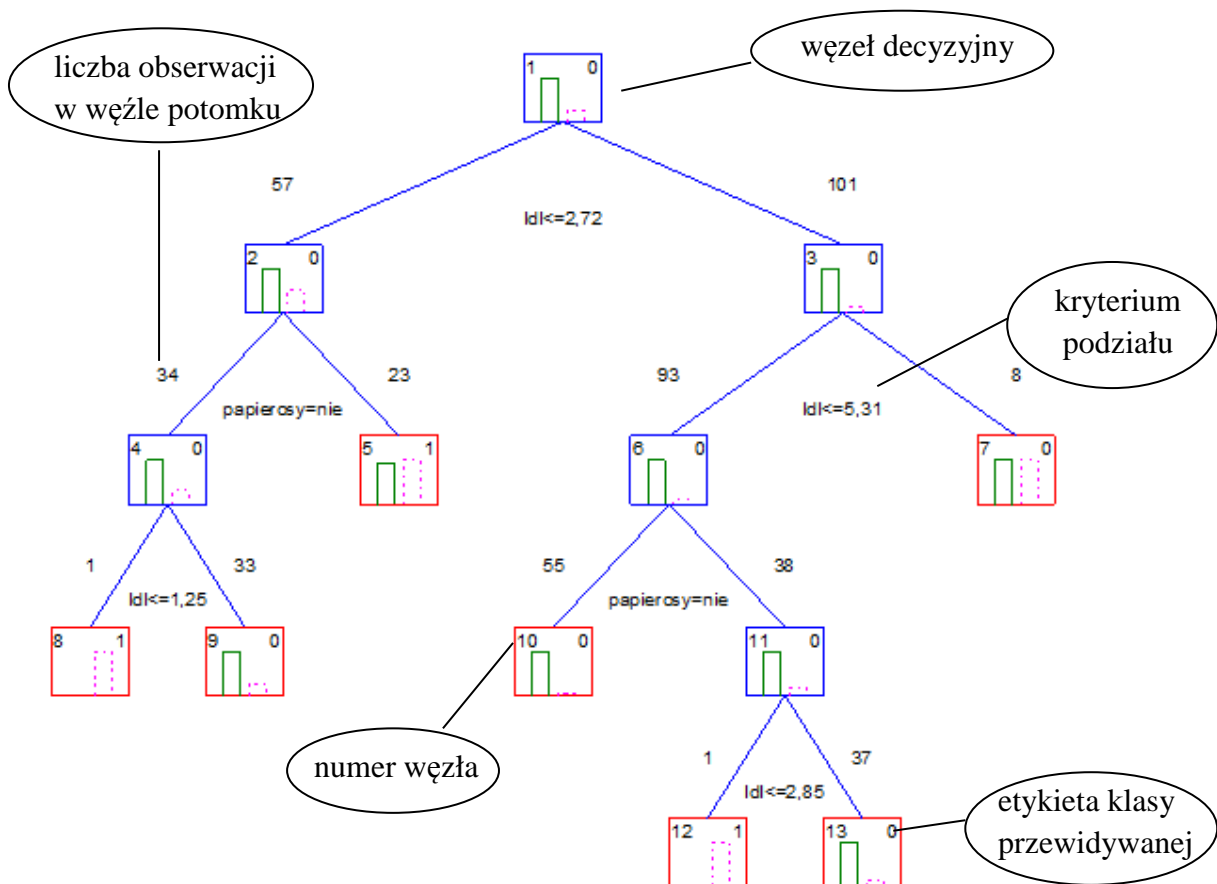
a) definicja drzewa klasyfikacyjnego

Drzewem klasyfikacyjnym nazywa się acykliczny spójny graf skierowany. Acykliczność oznacza, że nie istnieje ciąg krawędzi łączący którykolwiek wierzchołek z nim samym, natomiast spójność mówi o tym, że każda para wierzchołków jest połączona ciągiem krawędzi. Krawędzie takiego grafu nazywane są *gałęziami*. Wierzchołki, z których wychodzi co najmniej jedna krawędź nazywamy *węzłami*. Wierzchołki nie będące węzłami nazywamy *liśćmi*.



Wykres 5 Budowa drzewa klasyfikacyjnego, opracowanie własne

Każdemu węzłowi, będącemu liściem, przyporządkowane jest oznaczenie klasy, a każdej z gałęzi reguła decyzyjna, czyli warunek odnoszący się do wartości zmiennych w zbiorze wejściowym i mówiący, w jakim przypadku należy pójść daną gałęzią [133].



Wykres 6 Przykład drzewa klasyfikacyjnego, z uwzględnieniem kryteriów podziału, liczby obserwacji w węzłach oraz etykiet klas przewidywanych, opracowanie własne

W drzewie klasyfikacyjnym wewnętrzne węzły opisują, w jaki sposób dokonano podziału na jednorodne klasy (odpowiadają testom przeprowadzonym na wartościach atrybutów). Podział ten jest dokonywany w oparciu o wartości cech obiektów. Liście drzewa są odpowiednikami etykiet klas danego problemu dyskryminacji. Gałęzie drzewa reprezentują wartości cech, na podstawie których dokonano podziału (wyniki testów).

Drzewo klasyfikacyjne „rośnie” od korzenia do liści. Elementy próby przechodzą przez węzły wzdłuż gałęzi. W węzłach dokonuje się podział elementów na podgrupy, zgodnie z kryteriami podziału w nich zapisanych. Na tej podstawie następuje wybór gałęzi, wzdłuż której następuje dalszy podział, aż do momentu otrzymania liścia z etykietą którejś z klas. Dla każdego liścia istnieje tylko jedna droga łącząca go z korzeniem. Możliwa jest konwersja drzewa decyzyjnego do zbioru reguł klasyfikujących przykłady, która wykorzystywana jest przy przycinaniu drzewa zapobiegającemu nadmiernemu dopasowaniu.

b) metody konstrukcji drzew

Celem jest zbudowanie drzewa klasyfikacyjnego jak najmniejszego (o możliwie jak najmniejszej liczbie węzłów), tak, by otrzymane reguły klasyfikacji były jak najprostsze. Z drugiej strony należy skonstruować drzewo o jak najmniejszym błędzie rzeczywistym, czyli błędzie, który powstaje podczas testowania drzewa na elementach nienależących do zbioru przykładów.

Budowa drzewa klasyfikacyjnego przebiega w czterech powiązanych ze sobą etapach:

- określenie kryterium trafności przewidywania;
- wybór metody „optymalnego” podziału węzłów;
- ustalenie reguły zatrzymania podziału i reguła decyzyjna;
- wybór właściwej wielkości drzewa.

🚧 określenie kryterium trafności przewidywania

Drzewa klasyfikacyjne buduje się po to, by potem móc klasyfikować nowe obserwacje, o których nie mamy informacji o przynależności klasowej. W związku z tym celem jest uzyskanie najbardziej trafnego przewidywania. Za najbardziej trafne przewidywanie uznaje się to, które wymaga najmniejszych *kosztów*, czyli takie, które charakteryzuje się najmniejszym odsetkiem błędnych klasyfikacji.

Niekiedy pojawiają się sytuacje, w których niektóre błędne przewidywania są bardziej szkodliwe niż inne lub w których niektóre błędne prognozy występują częściej niż inne. W takich przypadkach należy bardziej zwrócić uwagę na potrzebę minimalizacji kosztów niż tylko na proporcję błędnie zaklasyfikowanych przypadków.

W niektórych sytuacjach oczekuje się jest trafniejszej klasyfikacji w przypadku pewnych klas niż w przypadku innych, bez względu na wielkości klas (np. bardziej pożądane będzie rozpoznanie nosiciela zarażającego od nosiciela niezarażającego), co powinno znaleźć swoje odzwierciedlenie w ustaleniu wyższych kosztów błędnej klasyfikacji.

Jedną z metod pozwalających na uzyskanie najbardziej trafnego przewidywania opiera się na wykorzystaniu prawdopodobieństwa *a priori*. Prawdopodobieństwa *a priori* określają, na ile jest prawdopodobne, bez żadnej wstępnej wiedzy na temat wartości zmiennych predykcyjnych w modelu, że dany przypadek lub obiekt należy do danej klasy. Względne wielkości tych prawdopodobieństw przypisanych do każdej klasy są wykorzystywane do skorygowania "ważności" błędnych klasyfikacji dla każdej klasy [134]. W przypadku ustalenia prawdopodobieństwa *a priori* proporcjonalnie do wielkości klas i równych dla każdej klasy kosztów błędnej klasyfikacji, minimalizacja kosztów jest równoznaczna z minimalizacją ogólnej proporcji przypadków zaklasyfikowanych błędnie.

Kolejną metodą służącą minimalizacji kosztów jest wykorzystanie wag przypadków zakodowanych w zmiennej ważącej. Alternatywą dla stosowania wag jest ustalenie odpowiednich prawdopodobieństw *a priori* i kosztów błędnej klasyfikacji.

Minimalizacja kosztów jest podstawowym celem analizy drzew klasyfikacyjnych. Jako drzewo „właściwej wielkości” przyjmuje się to, dla którego uzyskano minimalne oszacowane koszty.

wybór metody „optymalnego” podziału węzłów

Budowa drzewa decyzyjnego rozpoczyna się od odpowiedzi na pytanie, w jaki sposób wybrać zmienną, która najlepiej przyporządkowuje obiekty do n klas zmiennej zależnej. Wybór atrybutu i ustalenie dla niego kryterium podziału zależy od rodzaju budowanego drzewa.

Najprostszy podział drzew klasyfikacyjnych to podział na:

- drzewa binarne – to drzewa, w których z każdego węzła wychodzą jedynie dwie krawędzie;
- drzewa niebinarne – w których z węzła mogą wychodzić więcej niż dwie krawędzie.

W celu wyboru zmiennej predykcyjnej służącej do dokonania podziału w wybranym węźle, dla każdego węzła końcowego oblicza się poziomy prawdopodobieństwa testowego p dla testów istotności zależności przynależności do klasy od poziomów każdej zmiennej predykcyjnej. Zmienna predykcyjna, dla której występuje najmniejszy poziom prawdopodobieństwa p (mniejszy od poziomu istotności skorygowanego zgodnie z nierównością Bonferroniego dla porównań wielokrotnych), zostaje wybrana do podziału odpowiadającego jej węzła [134].

Liczba możliwych kryteriów podziału dla zmiennej niezależnej zależy od jej typu [135]:

- dla zmiennej binarnej – jest tylko jeden możliwy podział;
- dla zmiennej nominalnej z k poziomami – jest $2^{(k-1)} - 1$ możliwych kontrastów między dwoma zbiorami poziomów tego predyktora, co wiąże się z wykonaniem $2^{(k-1)} - 1$ porównań;
- dla zmiennej ciągłej X mającej n różnych wartości – należy wykonać $n - 1$ porównań, co wynika z tego, że sortując rosnąco wartości predyktora $x_1 \leq x_2 \leq \dots \leq x_n$, kryteria podziału definiuje się jako $X \leq \text{średnia}(x_k, x_{k+1})$ (przy danym węźle mamy $k - 1$ środkowych punktów między różnymi poziomami).

Miary dobroci dopasowania

W węźle następuje podział podzbioru na części. Celem jest znalezienie takiego kryterium podziału, by uzyskać jak najmniejszą różnorodność klas w tych częściach. Optymalnym

byłoby osiągnięcie możliwie dużej różnicy pomiędzy różnorodnością klas w węzle i różnorodnością klas w tych częściach.

Dla każdego węzła t można określić miarę niejednorodności (różnorodności) w następujący sposób:

$$i(t) = \phi(p(1|t), \dots, p(K|t)),$$

gdzie ϕ oznacza funkcję określoną dla wszystkich K -elementowych ciągów prawdopodobieństw (p_1, p_2, \dots, p_K) takich, że $\sum_{i=1}^K p_i = 1$, $p_i \geq 0$ ($i=1, 2, \dots, K$) spełniającą następujące warunki:

- funkcja ϕ przyjmuje maksimum tylko w punkcie $(\frac{1}{K}, \frac{1}{K}, \dots, \frac{1}{K})$;
- funkcja ϕ przyjmuje minimum tylko w punktach: $(1, 0, \dots, 0)$, $(0, 1, \dots, 0)$, \dots , $(0, 0, \dots, 1)$;
- funkcja ϕ jest symetryczną funkcją swoich argumentów.

Najbardziej powszechne miary stosowane w celu wyboru najlepszego podziału to [135]:

- miara błędnych klasyfikacji, obliczana w następujący sposób

$$\phi(p_1, p_2, \dots, p_K) = 1 - \max\{p_1, p_2, \dots, p_K\}$$

- kryterium wykorzystujące test niezależności χ^2 – stosowane wtedy, gdy zmienne niezależne są jakościowe lub dyskretne; na każdym etapie tworzenia drzewa buduje się tabelę kontyngencji. Jeśli zmienna zależna ma $a > 2$, a zmienna niezależna $b > 2$ kategorii, wówczas tabelę kontyngencji o wymiarach $a \times b$ redukuje się do mniejszej poprzez łączenie w dozwolony sposób kategorii zmiennej niezależnej;
- kryterium Giniego definiowane wzorem

$$\phi(p_1, p_2, \dots, p_K) = 1 - \sum_{i=1}^K p_i^2$$

- reguła podziału na dwie części (Twoing criterion) – stosowana dla każdego typu zmiennych niezależnych w przypadku, gdy zmienna zależna ma przynajmniej 3 kategorie; jest wykorzystywana, gdy pożądana jest zamiana poszukiwań optymalnego podziału na k kategorii na optymalny podział na 2 superklasy;
- uporządkowana reguła podziału na dwie części (ordered Twoing criterion) – używana wtedy, gdy zmienna zależna posiada co najmniej 3 uporządkowane kategorie, jest to reguła, w której dwie superklasy zawierają tylko następujące po sobie („sąsiadujące”) kategorie;
- entropia

$$\phi(p_1, p_2, \dots, p_K) = -\sum_{i=1}^K p_i \log p_i.$$

Wybierany jest podział, w którym otrzymuje się maksymalną redukcję niejednorodności przynależności do klasy w węźle.

ustalenie reguły zatrzymania podziału i reguła decyzyjna

Konstrukcja drzewa rozpoczyna się od jednego wierzchołka (korzenia), któremu został przypisany zbiór uczący oraz zbiór dostępnych reguł (testów). Wzrost drzewa powoduje zmniejszenie zarówno liczby elementów w zbiorze uczącym jak i dostępnych testów. W przypadku budowy zbyt dużego drzewa klasyfikacyjnego pojawia się problem z tzw. efektem przetrenowania, który polega na tym, że drzewo doskonale klasyfikuje elementy z próby uczącej, jednakże coraz słabiej, w miarę zwiększania liczby liści, nowe obiekty.

Pojawia się zatem problem budowy drzewa o optymalnej wielkości. Jednym z rozwiązań tego zagadnienia jest zastosowanie tzw. reguły stopu. Jeden z tego typu warunków opiera się na minimalnej liczności w węzłach. Według tej zasady, podziały mogą zostać zakończone, gdy wszystkie węzły końcowe są czyste (nie zawierają błędnych klasyfikacji) lub zawierają nie więcej niż określoną minimalną liczbę obiektów. Niekiedy stosuje się modyfikację tego warunku polegającą na tym, że podziały są kontynuowane do czasu, gdy wszystkie węzły

końcowe są czyste lub zawierają nie więcej przypadków niż określona minimalna frakcja licznosci jednej lub więcej klas.

Według innej reguły stopu, należy zakończyć dalszy podział w wierzchołku, jeśli:

- w danym wierzchołku otrzymanie klasy k jest zdarzeniem z prawdopodobieństwem warunkowym równym 1;
- zastosowanie dowolnego, dostępnego podziału daje zerową lub ujemną zmianę miary jednorodności klas;
- zostały już wykorzystane wszystkie reguły.

Kiedy zostanie zakończony podział, ostatnie wierzchołki stają się liśćmi. Określona na liściach drzewa funkcja ind , definiowana jako

$$ind(t) = \arg \max_{1 \leq i \leq K} p(i | t)$$

przyporządkowuje etykietę klasy każdemu z węzłów końcowych.

Drzewo uznajemy za zbudowane, jeżeli, według reguły stopu, nie istnieją wierzchołki, w których należy dokonać dalszych podziałów.

🚧 wybór właściwej wielkości drzewa

Problem nadmiernego dopasowania

Mały błąd klasyfikacji uzyskany na próbie uczącej i duży błąd rzeczywisty, mogą świadczyć o nadmiernym dopasowaniu drzewa do danych trenujących. Błąd rzeczywisty nadmiernie dopasowanych drzew można zmniejszyć poprzez ich uproszczenie zwane przycinaniem, które powoduje zmniejszenie struktury drzewa, co z jednej strony skraca czas klasyfikacji, ale z drugiej strony pogarsza jej dokładność.

Przycinanie drzewa klasyfikacyjnego

Proces przycinania polega na zastąpieniu drzewa wyjściowego jego poddrzewem w taki sposób, że niektóre poddrzewa drzewa wyjściowego zostają zamienione na liście. Liściom tym przyporządkowuje się etykietę większościowej kategorii wśród obserwacji związanych z „ucinanym” poddrzewem.

Do przycinania wykorzystuje się tzw. zbiór przycinania, który może pochodzić spoza próby uczącej (w przypadku, gdy zbiór uczący jest wystarczająco duży) lub może pokrywać się ze zbiorem uczącym. Zbiór przycinania odgrywa szczególną rolę w szacowaniu błędu rzeczywistego przyciętego drzewa.

Istnieje wiele sposobów przycinania drzew klasyfikacyjnych. Jeden z nich polega na budowaniu drzewa określonej wielkości, zdefiniowanej przez użytkownika na podstawie wiedzy z wcześniejszych badań. Do przycinania można zastosować jedną z następujących procedur [134]:

- sprawdzian krzyżowy na podstawie próby testowej - drzewo budowane jest na podstawie próby uczącej, a jego trafność przewidywania jest testowana przez zastosowanie go do przewidywania przynależności do klas w próbie testowej. Koszty w próbie testowej przewyższające koszty w próbie uczącej wskazują na słaby wynik i sugerują, że drzewo innej wielkości mogłoby dać lepszy rezultat. Próby uczącą i testową można utworzyć gromadząc dwa niezależne zbiory danych lub, gdy mamy dużą próbę, wyodrębniając losową proporcję przypadków.
- v – krotny sprawdzian krzyżowy – przydaje się wtedy, gdy nie można skonstruować próby testowej ze względu na zbyt małą liczbę próby uczącej. Wartość v określa liczbę, wyodrębnianych losowo z próby uczącej, podzbiorów, możliwie równych sobie. Drzewo jest obliczane v razy, przy czym za każdym razem opuszcza się jedną z podprób i wykorzystuje ją jako próbę testową w sprawdzaniu krzyżowym (czyli każda z podprób jest wykorzystywana $v-1$ razy jako próba ucząca i raz jako próba testowa). Koszty sprawdzianu krzyżowego oblicza się dla każdej z v prób testowych, a następnie uśrednia. Otrzymuje się w ten sposób v -krotną ocenę kosztów sprawdzianu krzyżowego.
- globalny sprawdzian krzyżowy – polega na tym, że cała analiza powtarzana jest określoną liczbę razy, przy czym każdorazowo eliminuje się część próby uczącej

równą 1/liczbę powtórzeń. Każda wyeliminowana część próby jest wykorzystywana jako próba testowa w sprawdzianie krzyżowym wybranego drzewa.

Do budowy drzewa właściwej wielkości można również zastosować dobrze udokumentowane, ustrukturywane procedury opracowane w 1984 r. przez Breimana i in. Polegają one na przycinaniu drzewa na podstawie minimalizacji kosztów i złożoności drzewa w sprawdzianie krzyżowym (przycinanie koszt – złożoność).

Koszty te obliczane są w miarę wzrostu drzewa, od podziału przy węźle źródłowym aż do momentu, w którym osiągnie maksymalną wielkość określoną przez minimalną licznosc (n). Dodanie do drzewa kolejnych podziałów skutkuje obliczeniem kosztów dla próby uczącej, zwanych kosztami resubstytucji. Zatem liczba podziałów drzewa odpowiada ciągowi ogólnie malejących kosztów, odzwierciedlających lepszą klasyfikację. Należy pamiętać o tym, że jednocześnie, w momencie dodania do drzewa kolejnego podziału, wykonywany jest także v-krotny sprawdzian krzyżowy, dla którego obliczane są koszty sprawdzianu krzyżowego. Koszty dla węzła źródłowego wyznacza się na podstawie v-krotnego sprawdzianu krzyżowego opierając się na oszacowanych kosztach sprawdzianu krzyżowego.

Dla każdego drzewa (również tego, zbudowanego tylko z jednego węzła źródłowego) obliczana jest wartość funkcji zdefiniowanej jako: koszty dla każdego drzewa dodać parametr złożoności razy wielkość drzewa. Jest to tzw. funkcja „kary”, która jest liniową kombinacją (malejących wraz ze wzrostem drzewa) kosztów oraz liniowo wzrastającej wielkości drzewa. Parametr złożoności początkowo przyjmuje wielkość 0. Następnie zwiększa się jego wartość aż do momentu, w którym wartość funkcji dla największego drzewa przekroczy wartość funkcji dla mniejszego drzewa, które, w następnym kroku, stanie się nowym, największym drzewem i dla którego cała procedura zostanie powtórzona. Proces powtarza się tak długo, dopóki największym drzewem zostanie węzeł źródłowy.

Ciąg największych drzew, uzyskany przy zastosowaniu powyższej procedury, ma następujące własności:

- są one zagnieżdżone, gdyż kolejno przycinane drzewa zawierają wszystkie węzły następnego w kolejności mniejszego drzewa;
- przy dochodzeniu do węzła źródłowego przycina się coraz mniej węzłów;
- dla każdej wielkości drzewa w sekwencji nie ma innego drzewa tej samej wielkości, które miałyby mniejsze koszty [134].

c) najczęściej stosowane algorytmy w tworzeniu drzew klasyfikacyjnych

ID3 – stworzony przez Quinlana w 1983 r., może być wykorzystywany do tworzenia dowolnego rodzaju drzewa, cechuje się prostotą, ale wymaga kompletnych danych.

Zakłada, że dane są danymi dyskretnymi;

C 4.5 – jest uogólnieniem algorytmu ID3 i rozwiązujący większość jego problemów (braki danych, dane ciągłe, możliwość przycinania drzew);

C&RT – (Classification and Regression Trees) zaprojektowany w 1984 r. przez Breimana, Friedmana, Olshena i Stone’a dla binarnych drzew decyzyjnych. Wykorzystuje miarę Giniego, entropię oraz regułę podziału na dwie części;

CHAID – pomysłodawcami byli naukowcy pracujący w SPSS Inc., powstał w 1993 r., może być stosowany do budowy dowolnego drzewa. Jest to algorytm AID (Automatic Interaction Detection), wykorzystujący test niezależności chi – kwadrat.

6. Zastosowane oprogramowanie

Akwizycja sygnałów uzyskanych w wyniku 5-minutowej rejestracji zapisu elektrokardiograficznego odbyła się z wykorzystaniem systemu komputerowego KARDIOGRAF KARDIO PC (®MEDEA Gliwice). Następnie zostały one poddane analizie przy użyciu wirtualnego narzędzia, utworzonego w środowisku LabVIEW 7.1 Express firmy National Instruments [31].

W celu wykonania analizy statystycznej zastosowano oprogramowanie Statistica 10 firmy StatSoft.

7. Wyniki

7.1 Analiza jednowymiarowa

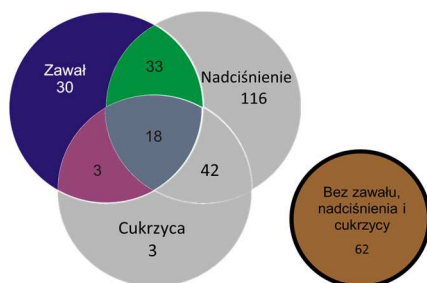
Wartości współczynników zmienności rytmu serca, wieku oraz BMI opisano za pomocą średniej arytmetycznej, odchylenia standardowego, mediany, minimum, maksimum oraz dolnego i górnego kwartyła. W analizie wstępnej dokonano porównań współczynników HRV, wieku oraz BMI. Ze względu na brak zgodności z rozkładem normalnym, zastosowano test U Manna-Whitneya.

Do zbadania zależności pomiędzy zmiennymi zmierzonymi na skali nominalnej użyto testu niezależności χ^2 .

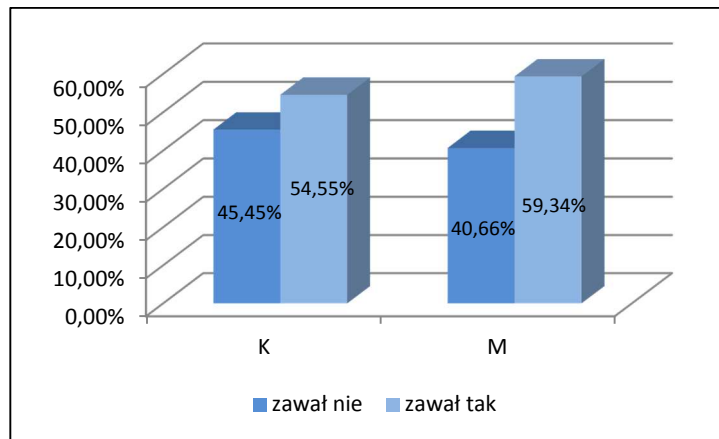
W badaniu:

- porównano współczynniki HRV, wiek oraz BMI pomiędzy **osobami** chorującymi na daną jednostkę chorobową i osobami z grupy kontrolnej oraz zbadano zależność pomiędzy płcią a zachorowalnością;
- porównano współczynniki HRV, wiek oraz BMI pomiędzy **kobietami** chorującymi na daną jednostkę chorobową i kobietami z grupy kontrolnej;
- porównano współczynniki HRV, wiek oraz BMI pomiędzy **mężczyznami** chorującymi na daną jednostkę chorobową i mężczyznami z grupy kontrolnej.

7.1.1 Pacjenci po przebytych zawałach mięśnia sercowego vs pacjenci z grupy kontrolnej



Za pomocą testu chi kwadrat stwierdzono, że zawał mięśnia sercowego tak samo często występuje w grupie kobiet i mężczyzn ($p=0,570$). Rozkład zachorowalności względem płci przedstawia poniższy wykres 7.

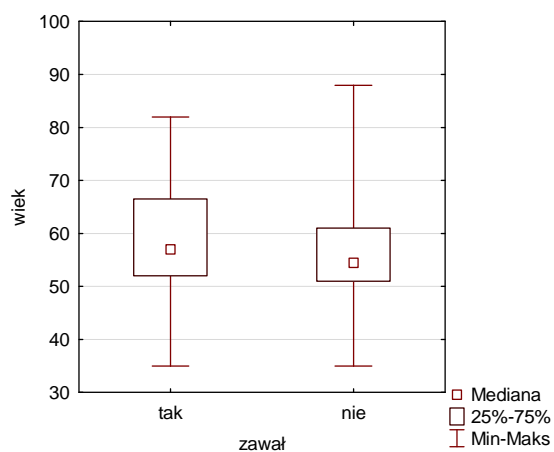


Wykres 7 Rozkład występowania zawału mięśnia sercowego względem płci, opracowanie własne

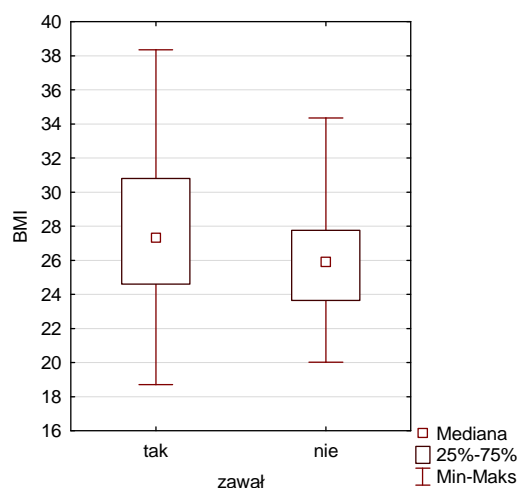
Nie wykazano istnienia różnic pomiędzy kobietami i mężczyznami po przebytym zawale mięśnia sercowego w wieku, wartościach współczynnika BMI oraz analizowanych wskaźnikach zmienności rytmu serca. Ponieważ zmienne nie miały rozkładu zgodnego z normalnym zastosowano do tego celu test U Manna-Whitneya.

Ze względu na brak zgodności z rozkładem normalnym analizowanych czynników (co zostało zbadane testem Shapiro-Wilka), dokonano porównań wartości parametrów zmienności rytmu serca, wieku i BMI z wykorzystaniem testu U Manna-Whitneya. Stwierdzono istnienie istotnych statystycznie różnic w wieku, BMI oraz w większości współczynników zmienności krótko i długoterminowej zmierzonych w dziedzinie czasu. Nie zaobserwowano istnienia istotnych różnic w wartościach, mierzonego w dziedzinie częstotliwości, wskaźnika LF/HF ($p=0,983$) oraz wskaźnika Allana, będącego parametrem mierzonym w połączonych dziedzinach czasu i częstotliwości ($p=0,245$).

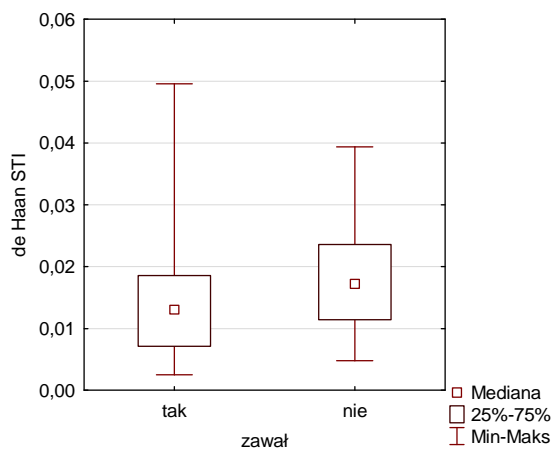
Osoby po przebytym zawale mięśnia sercowego były istotnie starsze od osób z grupy kontrolnej, miały wyższy wskaźnik BMI oraz istotnie niższe wartości wskaźników zmienności rytmu serca zmierzonych w dziedzinie czasu. Omawiane różnice ilustrują wykresy: wykres 8 – wykres 19. Szczegółowe dane dotyczące wartości poszczególnych parametrów oraz istotności statystycznej przedstawia **tabela A1 (Załącznik 1)**.



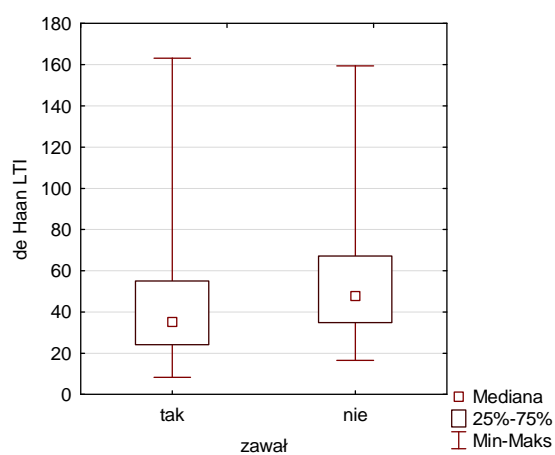
Wykres 8 rozkład wieku w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



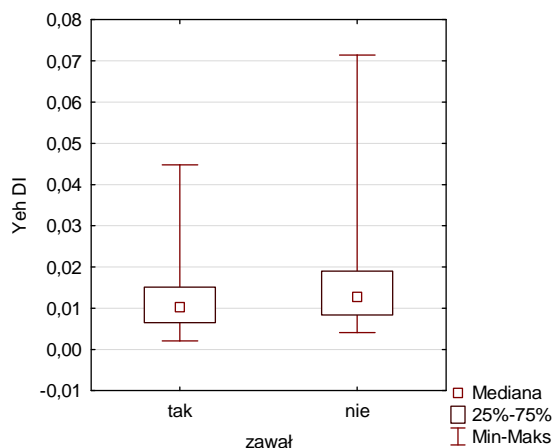
Wykres 9 rozkład BMI w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



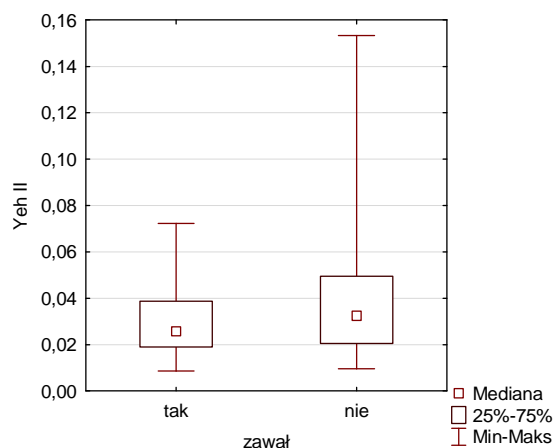
Wykres 10 rozkład wskaźnika zmienności krótkoterminowej de Haana w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



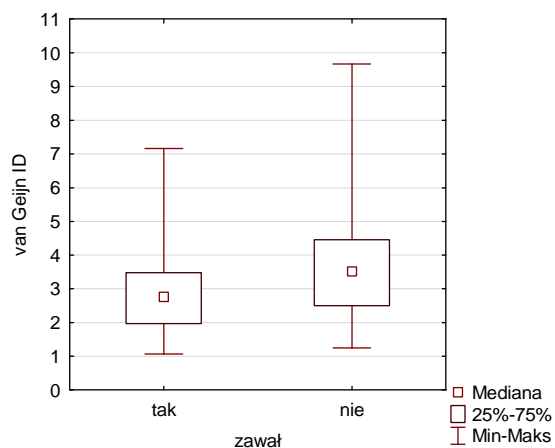
Wykres 11 rozkład wskaźnika zmienności długoterminowej de Haana w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



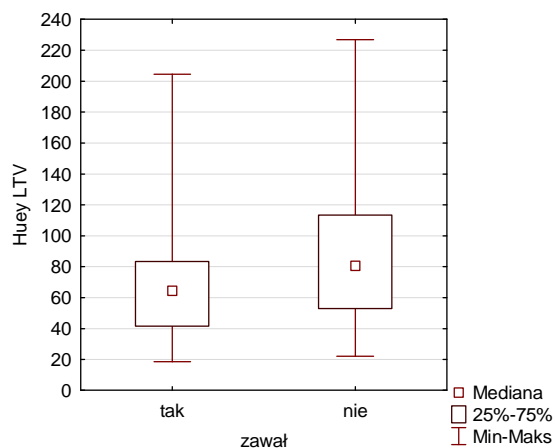
Wykres 12 rozkład wskaźnika zmienności krótkoterminowej Yeh'a w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



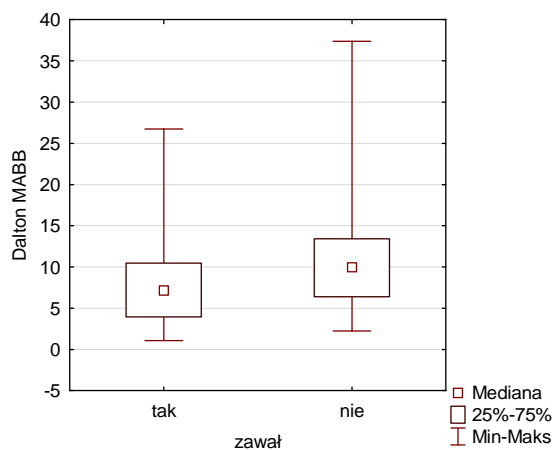
Wykres 13 rozkład wskaźnika zmienności długoterminowej Yeh'a w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



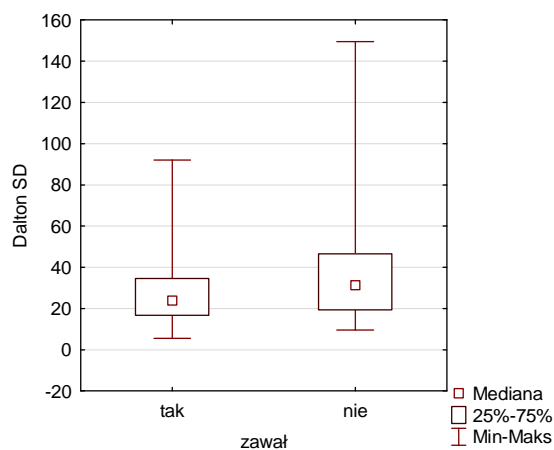
Wykres 14 rozkład wskaźnika zmienności krótkoterminowej van Geijna w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



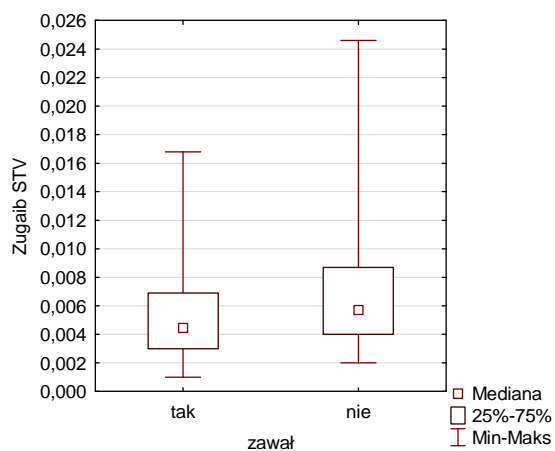
Wykres 15 rozkład wskaźnika zmienności długoterminowej Hueya w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



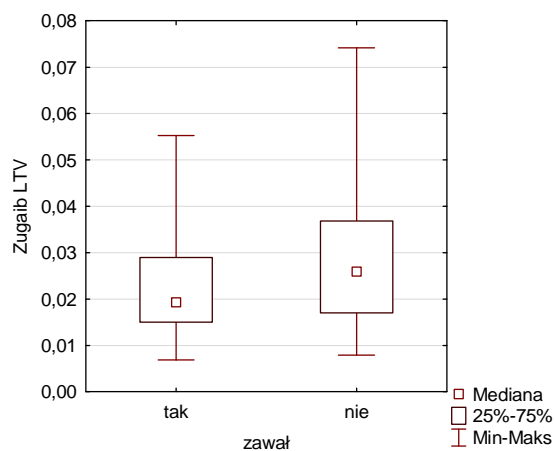
Wykres 16 rozkład wskaźnika zmienności krótkoterminowej Daltona w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



Wykres 17 rozkład wskaźnika zmienności długoterminowej Daltona w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



Wykres 18 rozkład wskaźnika zmienności krótkoterminowej Zugaiba w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej



Wykres 19 rozkład wskaźnika zmienności długoterminowej Zugaiba w grupie pacjentów po przebytych zawale mięśnia sercowego i w grupie kontrolnej

a) kobiety

W grupie kobiet, w celu zbadania istnienia istotnych statystycznie różnic pomiędzy wiekiem, BMI oraz wskaźnikami zmienności rytmu serca, z powodu braku zgodności z rozkładem normalnym, zastosowano test U Manna-Whitneya.

Stwierdzono istnienie istotnych statystycznie różnic wartościach wskaźnika BMI oraz w większości współczynników zmienności krótko i długoterminowej zmierzonych w dziedzinie czasu. Nie zaobserwowano istnienia istotnych różnic w wartościach, mierzonego w dziedzinie częstotliwości, wskaźnika LF/HF ($p=0,494$) oraz wskaźnika Allana, będącego parametrem mierzonym w połączonych dziedzinach czasu i częstotliwości ($p=0,132$). Kobiety, które przeszły zawał nie różniły się pod względem wieku od kobiet z grupy kontrolnej ($p=0,152$).

Kobiety po przebytych zawałach mięśnia sercowego miały wyższy wskaźnik BMI oraz istotnie niższe wartości wskaźników zmienności rytmu serca zmierzonych w dziedzinie czasu. Szczegółowe dane dotyczące wartości poszczególnych parametrów oraz istotności statystycznej przedstawia **tabela 14**.

Tabela 14 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród kobiet po przebytym zawale mięśnia sercowego i kobiet z grupy kontrolnej (test U Manna-Whitneya)

cecha	kobiety bez zawału, nadciśnienia i cukrzycy (N=25)				kobiety, które przeszły zawał (N=30)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
BMI	25,0 ± 3,2	25,2	20,0	31,6	27,9 ± 4,6	27,0	20,5	37,0	0,020
de Haan STI	0,020 ± 0,010	0,020	0,005	0,036	0,020 ± 0,010	0,010	0,003	0,050	0,043
de Haan LTI	54,4± 22,2	51,9	16,9	106,4	44,4 ± 34,9	33,3	8,3	163,2	0,020
Yeh II	0,040 ± 0,020	0,030	0,010	0,080	0,030 ± 0,020	0,020	0,009	0,070	0,022
van Geijn ID	4,14 ± 1,80	3,77	1,81	8,60	3,00 ± 1,30	3,00	1,07	6,30	0,019
Huey LTV	96,76 ± 45,00	94,20	27,23	185,86	67,40 ± 32,30	64,80	22,00	147,30	0,018
Dalton MABB	11,00 ± 6,40	10,30	2,30	36,40	8,32± 6,40	7,80	1,10	26,70	0,027
Dalton SD	34,50 ± 17,40	34,70	9,70	90,00	26,11 ± 17,80	21,30	5,70	92,00	0,023
Zugaib STV	0,007 ± 0,004	0,006	0,002	0,020	0,005 ± 0,003	0,005	0,001	0,013	0,032
Zugaib LTV	0,030 ± 0,010	0,030	0,008	0,050	0,020 ± 0,010	0,170	0,007	0,060	0,014
SDNN	18,10 ± 11,00	17,30	4,80	52,80	13,20 ± 8,80	11,10	2,80	45,50	0,026

b) mężczyźni

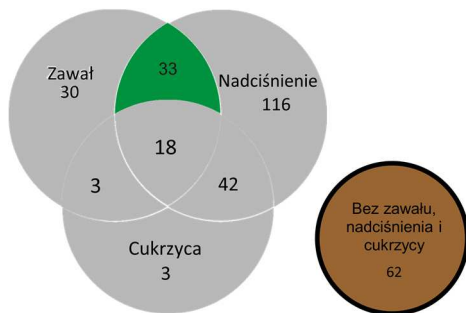
W grupie mężczyzn, w celu zbadania istnienia istotnych statystycznie różnic pomiędzy wiekiem, BMI oraz wskaźnikami zmienności rytmu serca, z powodu braku zgodności z rozkładem normalnym, zastosowano test U Manna-Whitneya.

Mężczyźni po przebytych zawale mięśnia sercowego różnili się istotnie statystycznie od mężczyzn z grupy kontrolnej względem wartości wskaźnika zmienności krótkoterminowej van Geijna ($p=0,046$). Mężczyźni po zawale mieli istotnie niższe wartości tego wskaźnika w porównaniu z mężczyznami z grupy kontrolnej. Szczegółowe dane dotyczące wartości poszczególnych parametrów oraz istotności statystycznej przedstawia **tabela 15**.

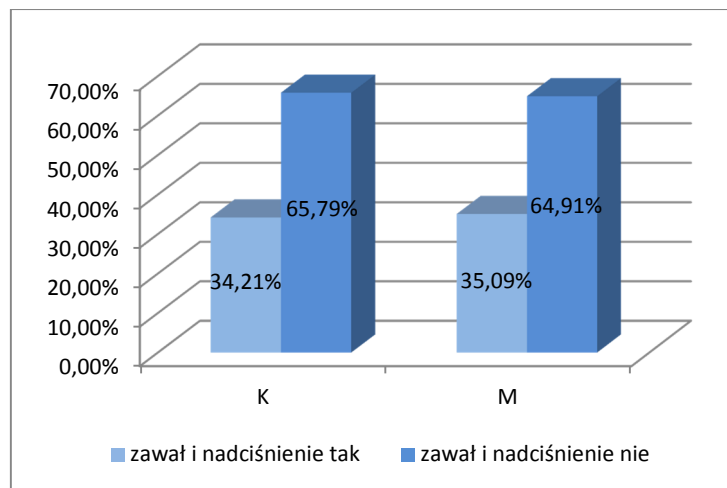
Tabela 15 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród mężczyzn po przebytym zawale mięśnia sercowego i mężczyzn z grupy kontrolnej (test U Manna-Whitneya)

cecha	mężczyźni bez zawału, nadciśnienia i cukrzycy (N=37)				mężczyźni, którzy przeszli zawał (N=54)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
van Geijn ID	3,60 ± 1,80	3,27	1,25	9,70	2,90 ± 1,40	2,70	1,12	7,20	0,046

7.1.2 Pacjenci po przebytych zawałach mięśnia sercowego i chorujący na nadciśnienie vs pacjenci z grupy kontrolnej



Za pomocą testu chi kwadrat stwierdzono, że płeć nie wpływa na częstość wystąpienia zawału mięśnia sercowego i zachorowalność na nadciśnienie ($p=0,930$). Rozkład zachorowalności względem płci przedstawia poniższy wykres 20.



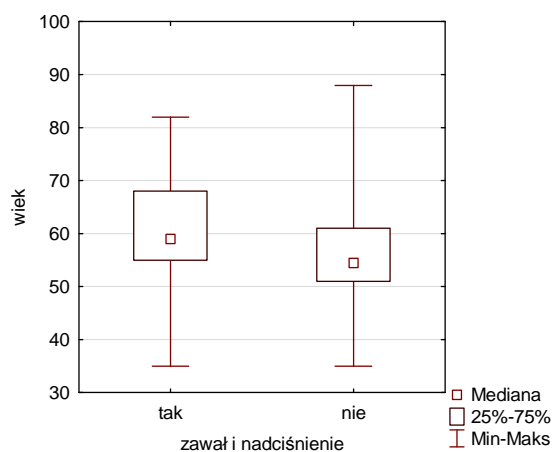
Wykres 20 Rozkład występowania zawału mięśnia sercowego i nadciśnienia tętniczego względem płci, opracowanie własne

Z powodu braku zgodności z rozkładem normalnym, w celu zbadania czy istnieją różnice pomiędzy kobietami i mężczyznami po przebytych zawałach mięśnia sercowego i chorującymi na nadciśnienie tętnicze w wieku, wartościach współczynnika BMI oraz analizowanych wskaźnikach zmienności rytmu serca, zastosowano test U Manna-Whitneya.

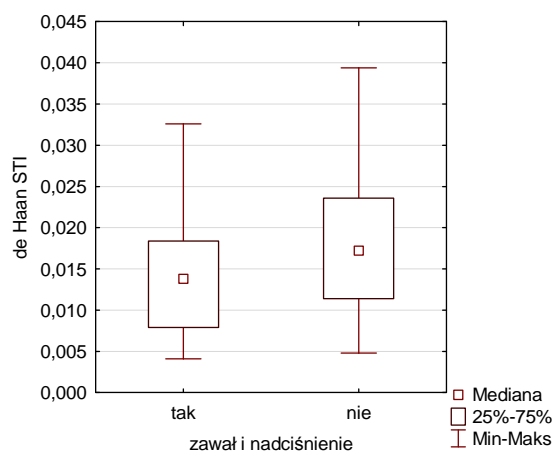
Ze względu na brak zgodności z rozkładem normalnym analizowanych czynników (co zostało zbadane testem Shapiro-Wilka), dokonano porównań wartości parametrów zmienności rytmu serca, wieku i BMI z wykorzystaniem testu U Manna-Whitneya. Stwierdzono istnienie istotnych statystycznie różnic w wieku, wartościach współczynników zmienności krótkoterminowej de Haana, van Geijna i Daltona i długoterminowej de Haana i Zugaiba, zmierzonych w dziedzinie czasu. Nie zaobserwowano istnienia istotnych różnic w

wartościach, mierzonego w dziedzinie częstotliwości, wskaźnika LF/HF ($p=0,218$), wskaźnika Allana, będącego parametrem mierzonym w połączonych dziedzinach czasu i częstotliwości ($p=0,524$) oraz BMI ($p=0,092$).

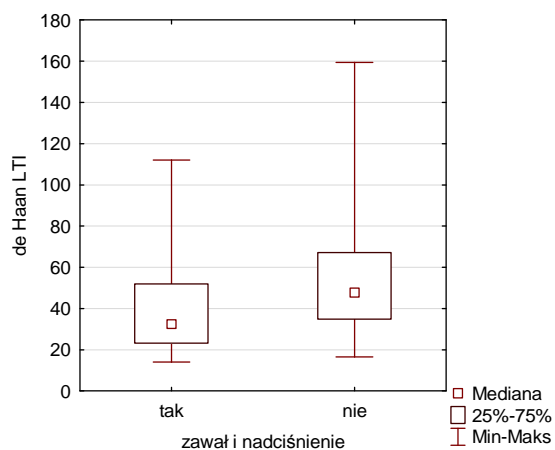
Osoby po przebytym zawale mięśnia sercowego dodatkowo chorujące na nadciśnienie tętnicze były istotnie starsze od osób z grupy kontrolnej oraz cechowały się istotnie niższymi wartościami wymienionych powyżej wskaźników zmienności rytmu serca. Omawiane różnice ilustrują wykresy: wykres 21 – wykres 26 . Szczegółowe dane dotyczące wartości poszczególnych parametrów oraz istotności statystycznej przedstawia **tabela A2 (Załącznik 1)**.



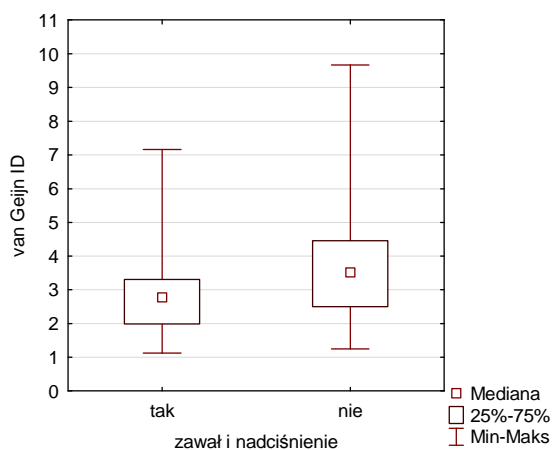
Wykres 21 rozkład wieku w grupie osób po przebytym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i w grupie kontrolnej



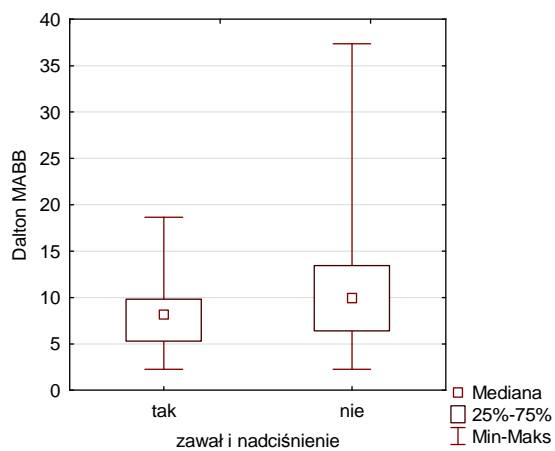
Wykres 22 rozkład wskaźnika zmienności krótkoterminowej de Haana w grupie osób po przebytym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i w grupie kontrolnej



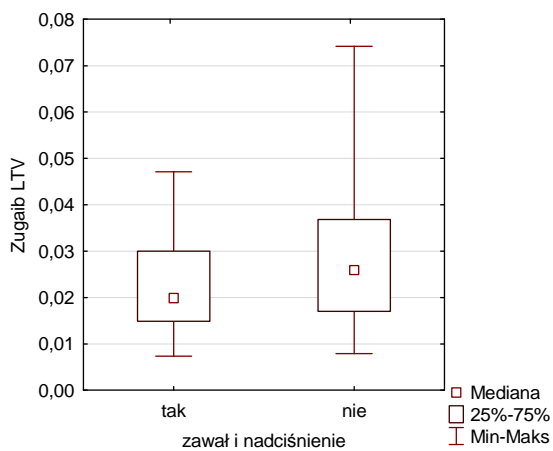
Wykres 23 rozkład wskaźnika zmienności długoterminowej de Haana w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i w grupie kontrolnej



Wykres 24 rozkład wskaźnika zmienności krótkoterminowej van Geijna w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i w grupie kontrolnej



Wykres 25 rozkład wskaźnika zmienności krótkoterminowej Daltona w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i w grupie kontrolnej



Wykres 26 rozkład wskaźnika zmienności długoterminowej Zugaiba w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i w grupie kontrolnej

a) kobiety

W grupie kobiet, które w przeszłości przebyły zawał i chorujących na nadciśnienie, w celu zbadania istnienia istotnych statystycznie różnic pomiędzy wiekiem, BMI oraz wskaźnikami zmienności rytmu serca w porównaniu do kobiet z grupy kontrolnej, z powodu braku zgodności z rozkładem normalnym, zastosowano test U Manna-Whitneya. Nie stwierdzono istnienia statystycznie istotnych różnic w wartościach analizowanych parametrów.

b) mężczyźni

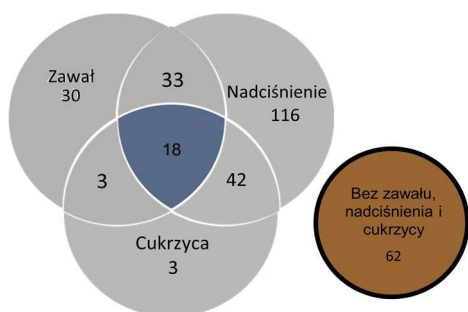
W grupie mężczyzn po przebyłym zawale mięśnia sercowego i chorujących na nadciśnienie, w celu zbadania istnienia istotnych statystycznie różnic pomiędzy wiekiem, BMI oraz wskaźnikami zmienności rytmu serca, w porównaniu z mężczyznami z grupy kontrolnej, z powodu braku zgodności z rozkładem normalnym, zastosowano test U Manna-Whitneya.

Mężczyźni po przebyłym zawale mięśnia sercowego cierpiący na nadciśnienie różnili się istotnie statystycznie od mężczyzn z grupy kontrolnej względem wartości wskaźnika zmienności długoterminowej de Haana ($p=0,007$). Mężczyźni po zawale chorujący na nadciśnienie mieli istotnie niższe wartości tego wskaźnika w porównaniu z mężczyznami z grupy kontrolnej. Szczegółowe dane dotyczące wartości wskaźnika de Hanna LTI oraz istotności statystycznej przedstawia **tabela 16**.

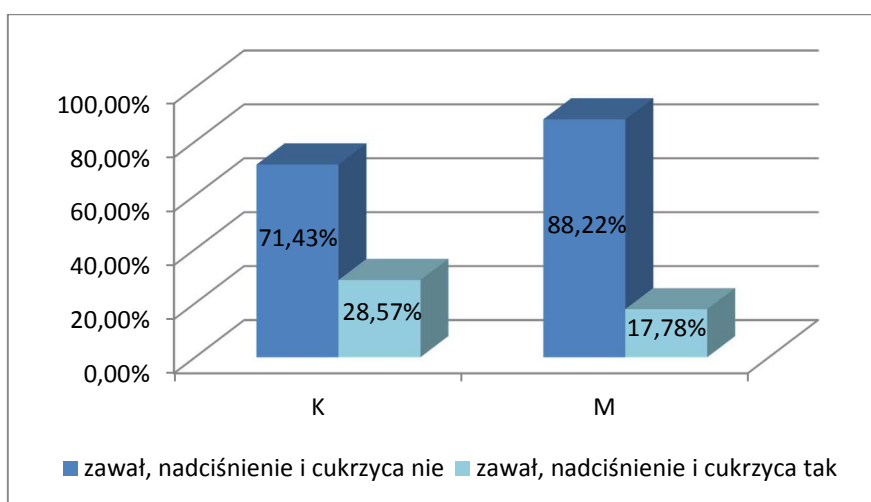
Tabela 16 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród mężczyzn po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze oraz mężczyzn z grupy kontrolnej (test U Manna-Whitneya)

cecha	mężczyźni bez zawału, nadciśnienia i cukrzycy (N=37)				mężczyźni, którzy chorują na zawał i nadciśnienie (N=20)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
de Haan LTI	55,2 ± 31,9	46,0	16,5	159,4	38,9 ± 22,3	30,4	15,7	96,3	0,007

7.1.3 Pacjenci po przebytych zawale mięśnia sercowego, którzy chorują na nadciśnienie i cukrzycę vs pacjenci z grupy kontrolnej



Obliczono test chi kwadrat w celu zbadania, czy zachorowalność na nadciśnienie i cukrzycę wraz z wystąpieniem zawału mięśnia sercowego zależy od płci. Nie stwierdzono istnienia takiej zależności ($p=0,251$). Rozkład zachorowalności względem płci przedstawia poniższy wykres 27.



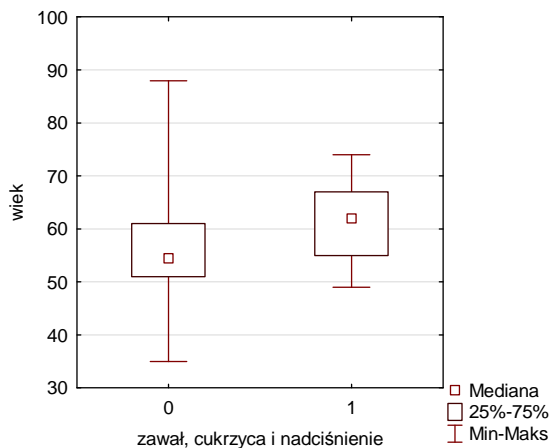
Wykres 27 Rozkład występowania zawału mięśnia sercowego, nadciśnienia tętniczego i cukrzycy względem płci, opracowanie własne

Nie wykazano istnienia różnic pomiędzy kobietami i mężczyznami po przebytych zawale mięśnia sercowego, chorującymi na nadciśnienie tętnicze i cukrzycę w wieku, wartościach współczynnika BMI oraz analizowanych wskaźnikach zmienności rytmu serca. Ponieważ zmienne nie miały rozkładu zgodnego z normalnym zastosowano do tego celu test U Manna-Whitneya.

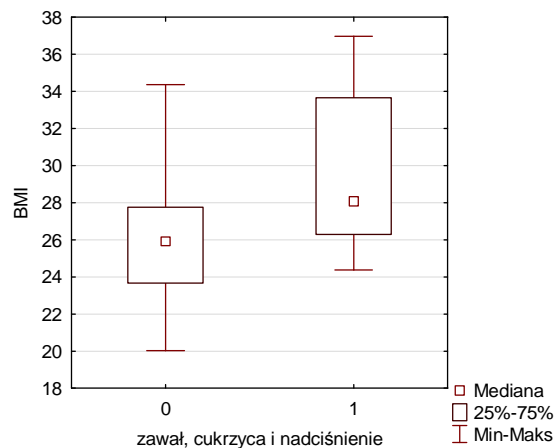
Dokonano porównań wartości parametrów zmienności rytmu serca, wieku i BMI z wykorzystaniem testu U Manna-Whitneya, ze względu na brak zgodności z rozkładem normalnym analizowanych czynników (co zostało zbadane testem Shapiro-Wilka).

Stwierdzono istnienie istotnych statystycznie różnic w wieku, wartościach współczynnika wagowo – wzrostowego BMI, w wartościach większości współczynników zmienności krótko i długoterminowej, zmierzonych w dziedzinie czasu oraz w wartościach parametru zmierzonego w połączonych dziedzinach czasu i częstotliwości – wskaźnika Allana. Nie zaobserwowano istnienia istotnych różnic w wartościach, mierzonego w dziedzinie częstotliwości, wskaźnika LF/HF ($p=0,071$) oraz zmierzonego w dziedzinie czasu wskaźnika zmienności długoterminowej Organa ($p=0,068$).

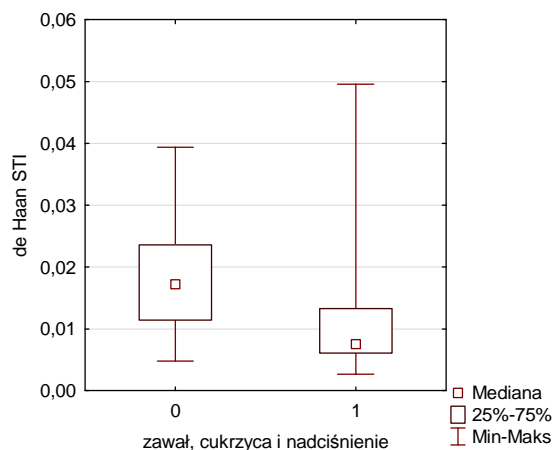
Osoby po przebytych zawale mięśnia sercowego dodatkowo chorujące na nadciśnienie tętnicze i cukrzycę były istotnie starsze od osób z grupy kontrolnej, miały istotnie wyższe wartości współczynnika BMI oraz cechowały się istotnie niższymi wartościami większości wskaźników zmienności rytmu serca zmierzonymi w dziedzinie czasu. Miały również istotnie wyższe wartości wskaźnika Allana. Omawiane różnice ilustrują wykresy: wykres 28 – wykres 44. Szczegółowe dane dotyczące wartości poszczególnych parametrów oraz istotności statystycznej przedstawia **tabela A3 (Załącznik 1)**.



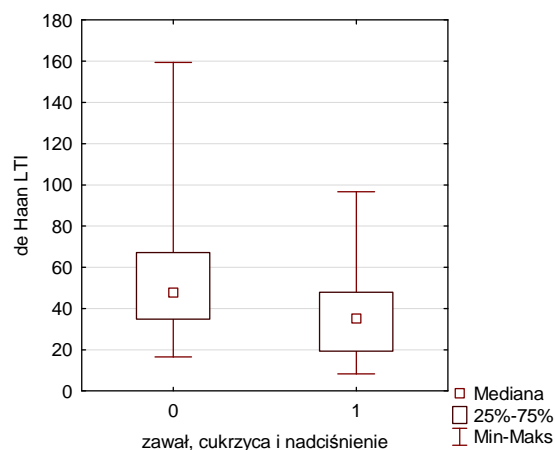
Wykres 28 rozkład wieku w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



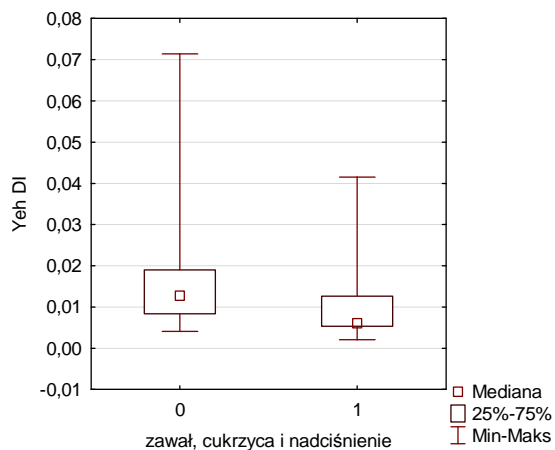
Wykres 29 rozkład wskaźnika BMI w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



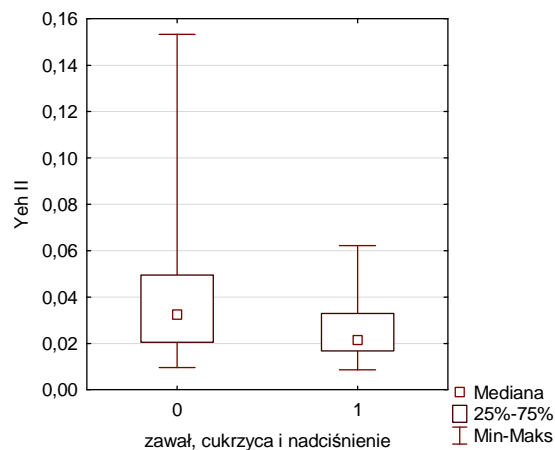
Wykres 30 rozkład wskaźnika zmienności krótkoterminowej de Haana w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



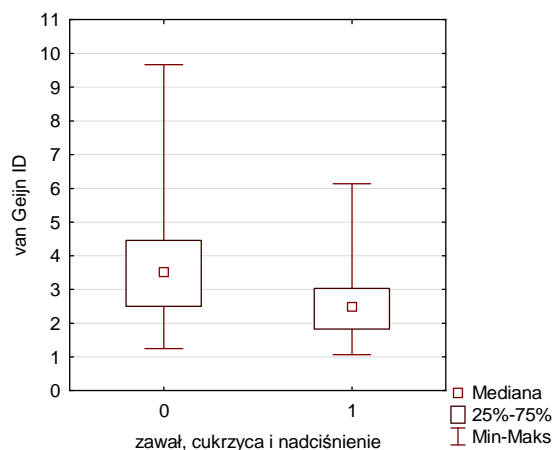
Wykres 31 rozkład wskaźnika zmienności długoterminowej de Haana w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



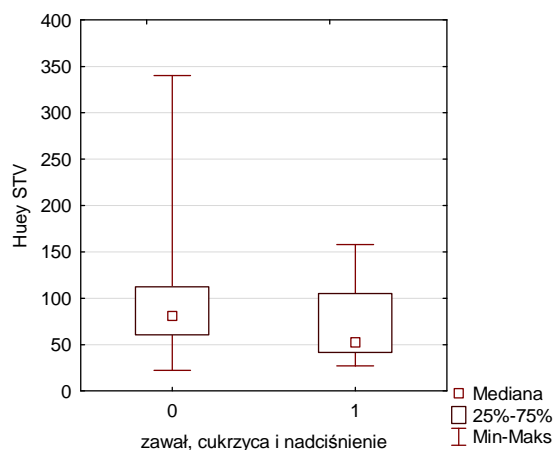
Wykres 32 rozkład wskaźnika zmienności krótkoterminowej Yeha w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



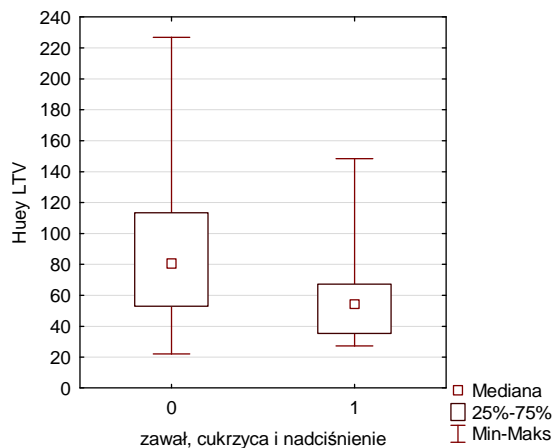
Wykres 33 rozkład wskaźnika zmienności długoterminowej Yeha w grupie osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



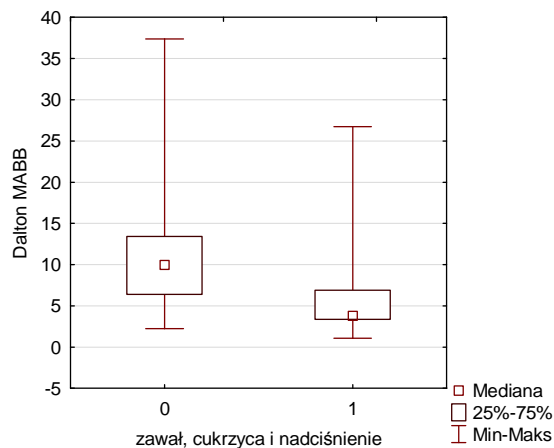
Wykres 34 rozkład wskaźnika zmienności krótkoterminowej van Geijna w grupie osób po przebyłym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



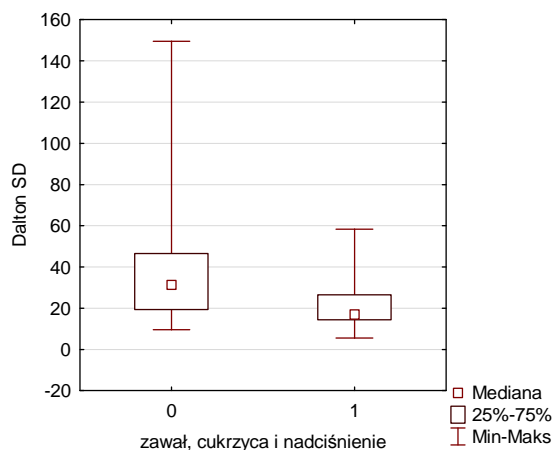
Wykres 35 rozkład wskaźnika zmienności krótkoterminowej Hueya w grupie osób po przebyłym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



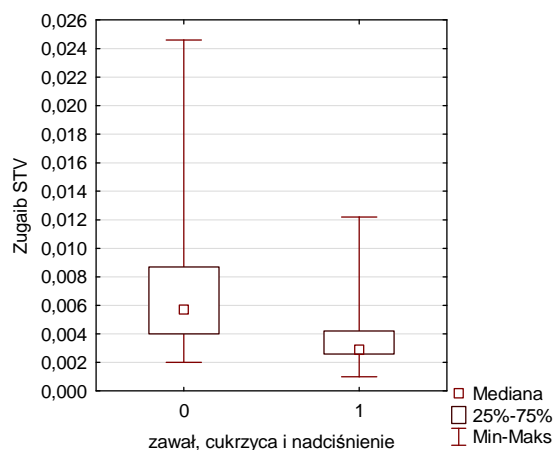
Wykres 36 rozkład wskaźnika zmienności długoterminowej Hueya w grupie osób po przebyłym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



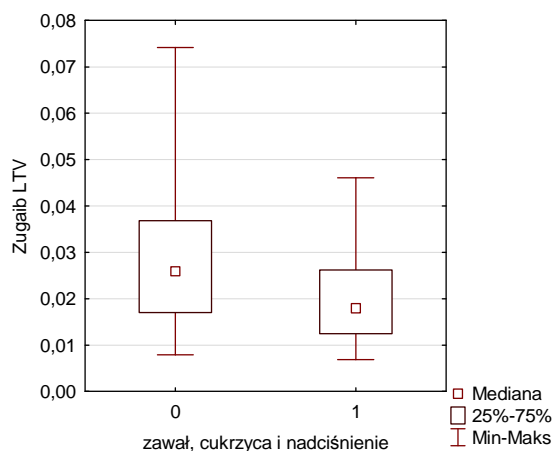
Wykres 37 rozkład wskaźnika zmienności krótkoterminowej Daltona w grupie osób po przebyłym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



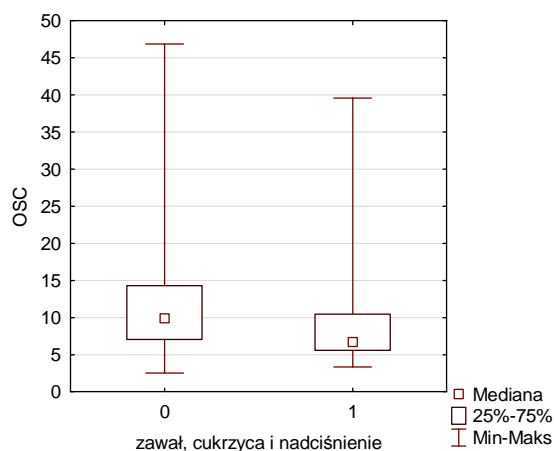
Wykres 38 rozkład wskaźnika zmienności długoterminowej Daltona w grupie osób po przebytych zawałach mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



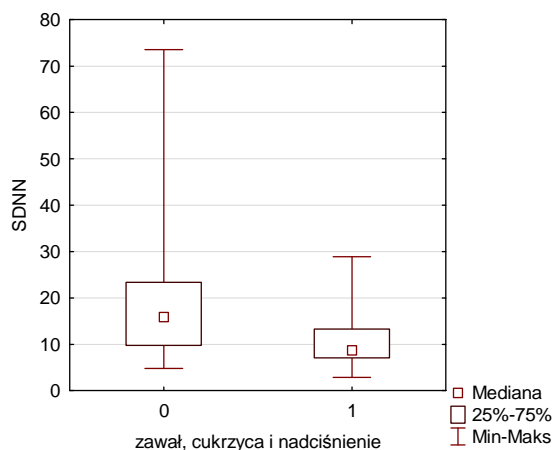
Wykres 39 rozkład wskaźnika zmienności krótkoterminowej Zugaiba w grupie osób po przebytych zawałach mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



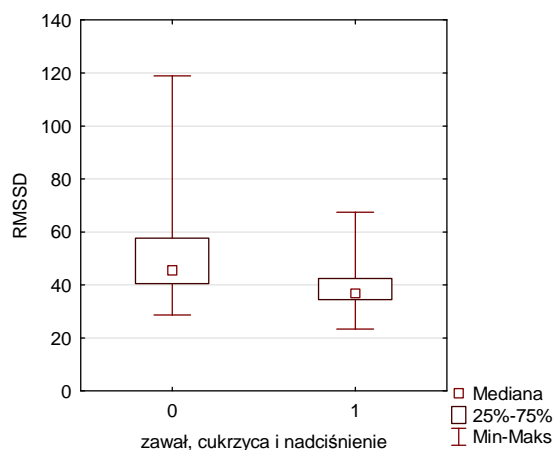
Wykres 40 rozkład wskaźnika zmienności długoterminowej Zugaiba w grupie osób po przebytych zawałach mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



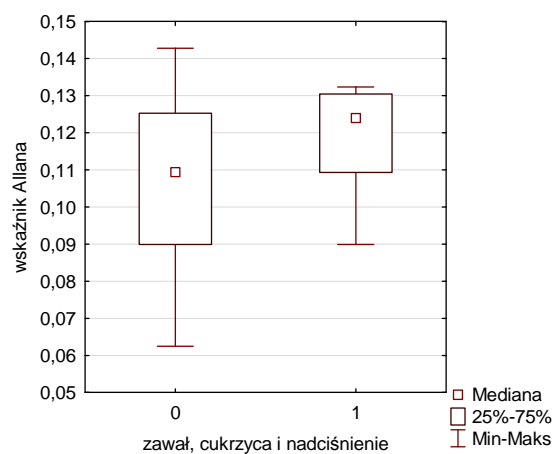
Wykres 41 rozkład wskaźnika szerokości wstęgi oscylacyjnej w grupie osób po przebytych zawałach mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



Wykres 42 rozkład wskaźnika SDNN w grupie osób po przebytym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



Wykres 43 rozkład wskaźnika RMSSD w grupie osób po przebytym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



Wykres 44 rozkład wskaźnika Allana w grupie osób po przebytym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)

a) kobiety

W celu sprawdzenia, czy istnieją statystycznie istotne różnice w wieku, BMI oraz wartościach wskaźników zmienności rytmu serca pomiędzy kobietami po przebytych zawale mięśnia sercowego, chorującymi dodatkowo na nadciśnienie tętnicze i kobietami z grupy kontrolnej, z powodu braku zgodności z rozkładem normalnym, zastosowano test U Manna-Whitneya.

Nie wykazano istnienia istotnych różnic w wieku ($p=0,439$). Kobiety po przebytych zawale, cierpiące na nadciśnienie tętnicze i cukrzycę nie różniły się od kobiet z grupy kontrolnej względem wskaźnika LF/HF, mierzonego w dziedzinie częstotliwości ($p=0,358$). Nie wykazano istnienia istotnych różnic w szerokości wstęgi oscylacyjnej ($p=0,225$), w wartościach wskaźnika zmienności krótkoterminowej Hueya ($p=0,303$) oraz Yeha ($p=0,065$), a także w wartościach parametru zmienności długoterminowej Organa ($p=0,141$). Kobiety po zawale mięśnia sercowego, chorujące na nadciśnienie tętnicze i cukrzycę mają istotnie wyższe wartości współczynnika BMI oraz istotnie niższe wartości wskaźników zmienności rytmu serca zmierzonych w dziedzinie czasu, zarówno krótko jak i długoterminowej, oraz wyższe wartości wskaźnika mierzonego w połączonych dziedzinach czasu i częstotliwości. Szczegółowe dane dotyczące wartości poszczególnych parametrów oraz istotności statystycznej przedstawia **tabela 17**.

Tabela 17 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród kobiet po przeżytym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę oraz kobiet z grupy kontrolnej (test U Manna-Whitneya)

cecha	Kobiety bez zawału, nadciśnienia i cukrzycy (N=25)				Kobiety, które chorują na zawał, nadciśnienie i cukrzycę (N=10)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
BMI	25,0±3,2	25,2	20,0	31,6	30,8±4,5	31,6	24,4	37,0	0,001
de Haan STI	0,019±0,008	0,019	0,005	0,036	0,013±0,014	0,008	0,003	0,050	0,005
de Haan LTI	54,4±22,2	51,9	16,9	106,4	36,1±26,5	32,6	8,3	96,6	0,021
Yeh II	0,038±0,017	0,034	0,010	0,076	0,026±0,016	0,021	0,009	0,062	0,034
van Geijn ID	4,14±1,85	3,77	1,81	8,61	2,79±1,45	2,70	1,07	6,14	0,038
Huey LTV	96,76±45,44	94,17	27,23	185,86	64,12±37,30	54,22	27,21	147,29	0,041
Dalton MABB	11,01±6,37	10,26	2,25	36,43	6,75±7,71	3,54	1,09	26,74	0,009
Dalton SD	34,69±17,42	34,70	9,68	90,01	21,37±14,69	16,33	5,66	53,32	0,025
Zugaib STV	0,007±0,004	0,006	0,002	0,020	0,005±0,004	0,003	0,001	0,012	0,019
Zugaib LTV	0,029±0,012	0,027	0,008	0,050	0,020±0,012	0,018	0,007	0,046	0,025
SDNN	18,15±10,91	17,26	4,81	52,80	10,70±7,30	8,37	2,85	26,66	0,023
RMSSD	50,1±15,0	45,5	28,7	101,0	36,9±10,7	34,8	23,4	64,7	0,001
wsk. Allana	0,11±0,02	0,11	0,06	0,14	0,12±0,01	0,13	0,10	0,13	0,038

b) mężczyźni

Zastosowano test U Manna-Whitneya, aby zbadać, czy istnieją statystycznie istotne różnice w wartościach wieku, współczynnika BMI oraz wskaźników zmienności rytmu serca pomiędzy mężczyznami z grupy kontrolnej a mężczyznami po przebytym zawale mięśnia sercowego i chorujących dodatkowo na nadciśnienie tętnicze i cukrzycę (z powodu braku zgodności rozkładów analizowanych zmiennych z rozkładem normalnym).

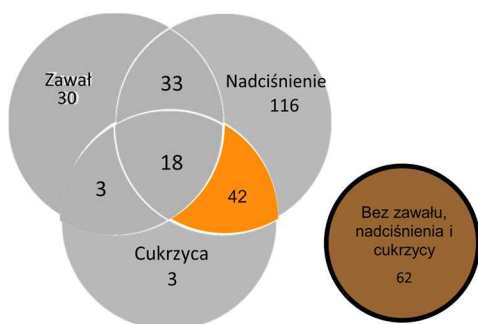
Mężczyźni po przebytym zawale mięśnia sercowego, cierpiący dodatkowo na nadciśnienie i cukrzycę byli istotnie starsi od mężczyzn z grupy kontrolnej ($p=0,012$). Nie stwierdzono istnienia istotnych różnic ani w wartościach współczynnika BMI ($p=0,126$), ani w wartościach analizowanych wskaźników zmienności rytmu serca.

Szczegółowe dane dotyczące wartości wieku oraz istotności statystycznej przedstawia **tabela 18**.

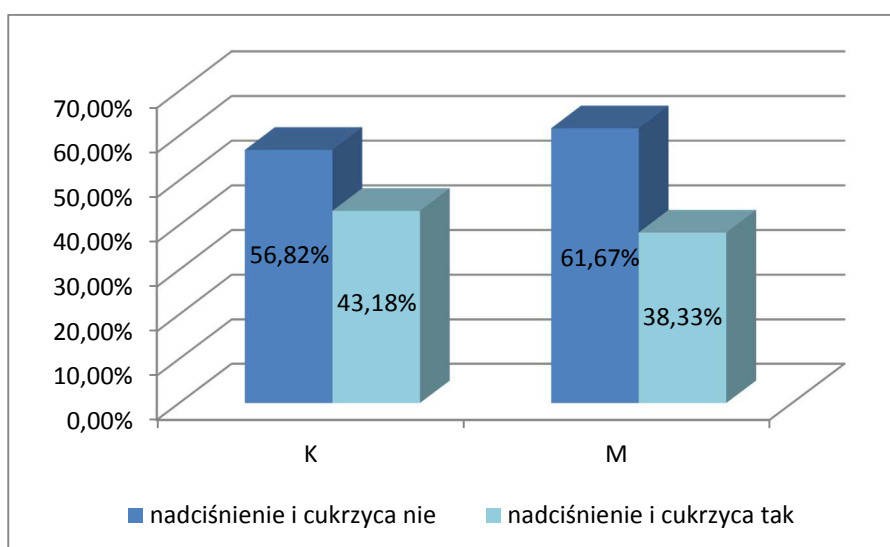
Tabela 18 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród mężczyzn po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze oraz cukrzycę i mężczyzn z grupy kontrolnej (test U Manna-Whitneya)

cecha	mężczyźni bez zawału, nadciśnienia i cukrzycy (N=37)				mężczyźni, którzy chorują na zawał, nadciśnienie i cukrzycę (N=8)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
wiek	54,8 ± 8,8	53,0	35,0	81,0	62,3 ± 6,5	64,0	54,0	72,0	0,007

7.1.4 Pacjenci, którzy chorują na nadciśnienie i cukrzycę vs pacjenci z grupy kontrolnej



W celu zbadania, czy zachorowalność na nadciśnienie i cukrzycę zależy od płci zastosowano test chi kwadrat. Stwierdzono, że nadciśnienie tętnicze i cukrzyca tak samo często występuje zarówno w grupie kobiet jak i mężczyzn ($p=0,619$). Rozkład wystąpienia nadciśnienia tętniczego łącznie z cukrzycą względem płci przedstawia poniższy wykres 45.



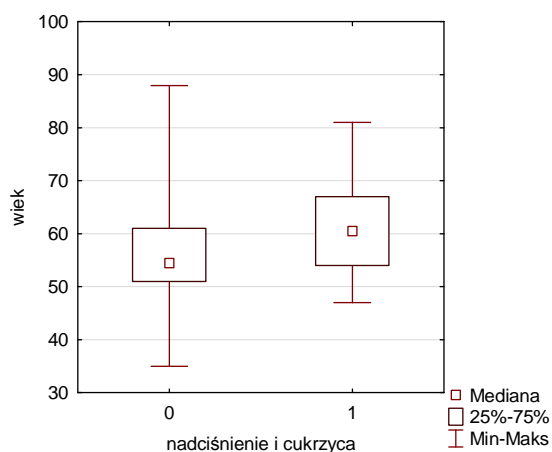
Wykres 45 Rozkład występowania nadciśnienia tętniczego i cukrzycy względem płci, opracowanie własne

Nie wykazano istnienia różnic pomiędzy kobietami i mężczyznami chorującymi na nadciśnienie tętnicze i cukrzycę w wieku, wartościach współczynnika BMI oraz analizowanych wskaźnikach zmienności rytmu serca. Z powodu braku zgodności z rozkładem normalnym, zastosowano do tego celu test U Manna-Whitneya.

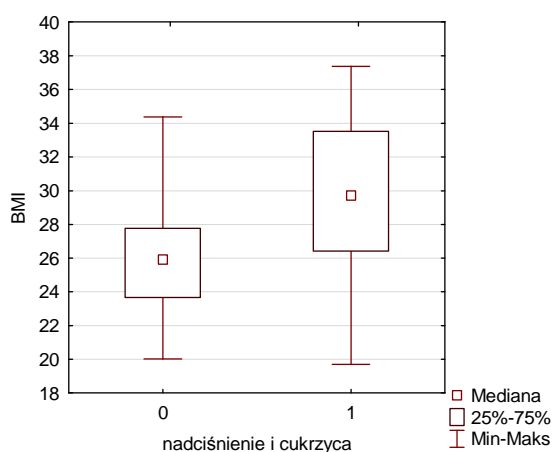
Dokonano porównań wartości parametrów zmienności rytmu serca, wieku i BMI z wykorzystaniem testu U Manna-Whitneya, ze względu na brak zgodności z rozkładem normalnym analizowanych czynników (co zostało zbadane testem Shapiro-Wilka). Osoby chorujące na nadciśnienie tętnicze i cukrzycę różniły się istotnie od osób z grupy kontrolnej względem wieku, wartości współczynnika wagowo – wzrostowego BMI, wartości

większości współczynników zmienności krótko i długoterminowej, zmierzonych w dziedzinie czasu oraz względem wartości parametru zmierzonego w połączonych dziedzinach czasu i częstotliwości – wskaźnika Allana.

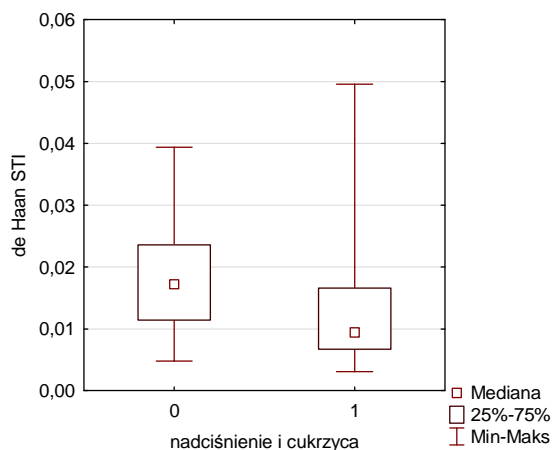
Osoby chorujące na nadciśnienie tętnicze i cukrzycę były istotnie starsze od osób z grupy kontrolnej, miały istotnie wyższe wartości współczynnika BMI oraz cechowały się istotnie niższymi wartościami większości wskaźników zmienności rytmu serca zmierzonymi w dziedzinie czasu. Miały również istotnie wyższe wartości wskaźnika Allana. Nie zaobserwowano istnienia istotnych różnic w wartościach, mierzonego w dziedzinie częstotliwości, wskaźnika LF/HF ($p=0,987$). Omawiane różnice ilustrują wykresy: wykres 46 – wykres 63. Szczegółowe dane dotyczące wartości poszczególnych parametrów oraz istotności statystycznej przedstawia **tabela A4 (Załącznik 1)**.



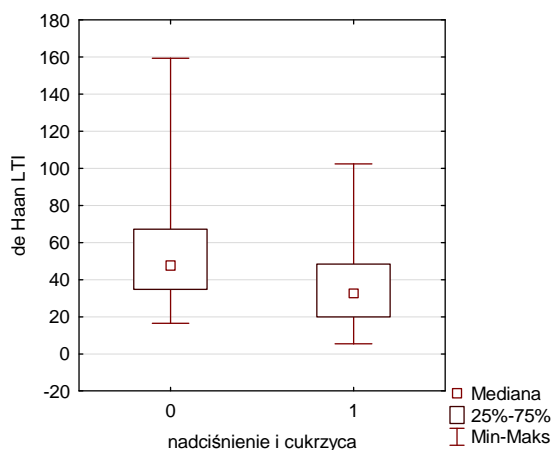
Wykres 46 rozkład wieku w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



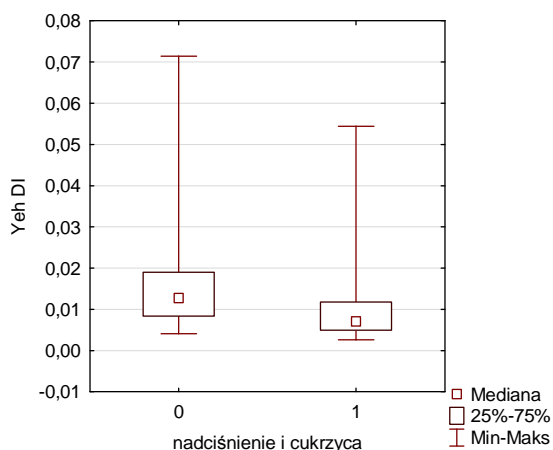
Wykres 47 rozkład wskaźnika BMI w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



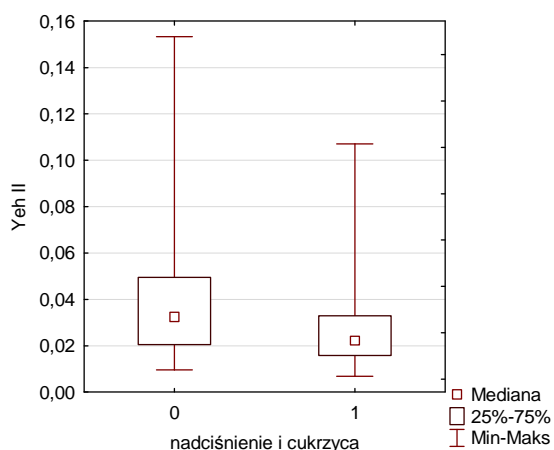
Wykres 48 rozkład wskaźnika zmienności krótkoterminowej de Haana w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



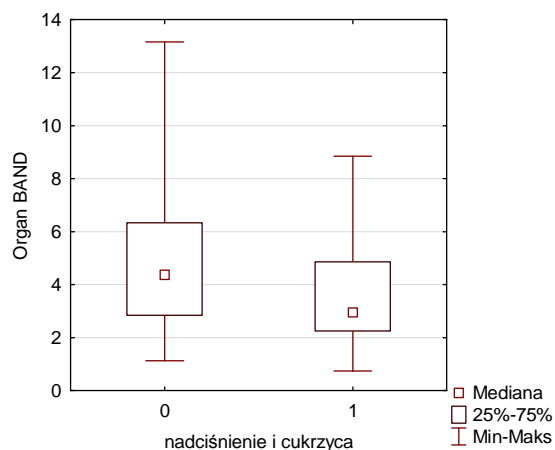
Wykres 49 rozkład wskaźnika zmienności długoterminowej de Haana w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



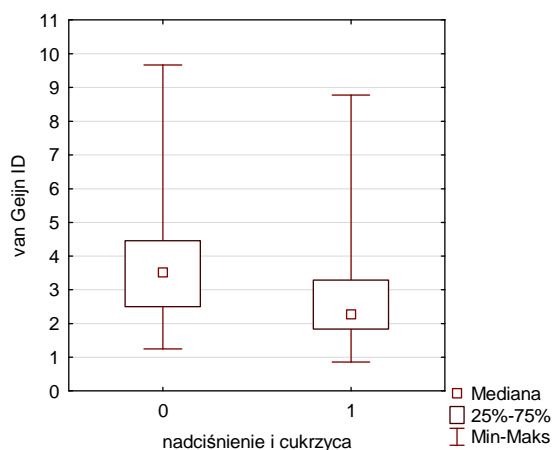
Wykres 50 rozkład wskaźnika zmienności krótkoterminowej Yeha w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



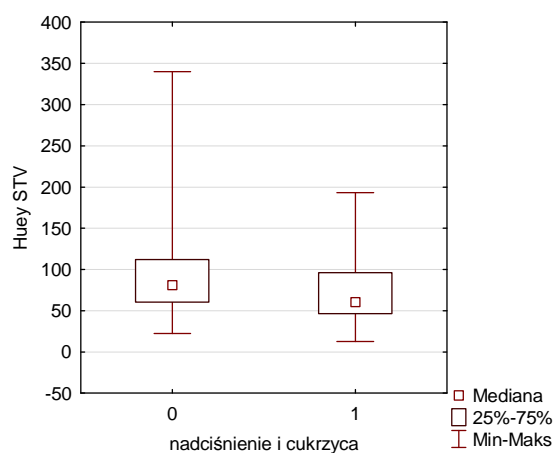
Wykres 51 rozkład wskaźnika zmienności długoterminowej Yeha w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



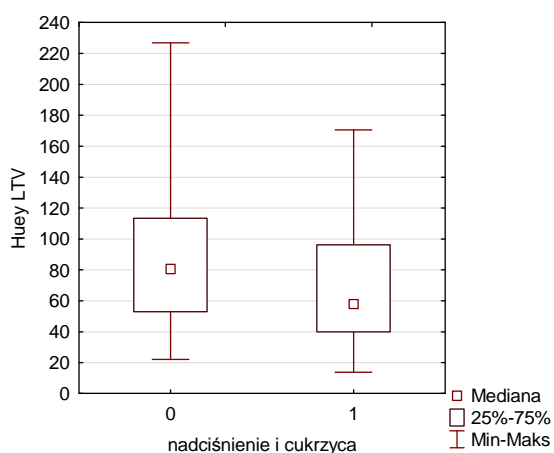
Wykres 52 rozkład wskaźnika zmienności długoterminowej Organa w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



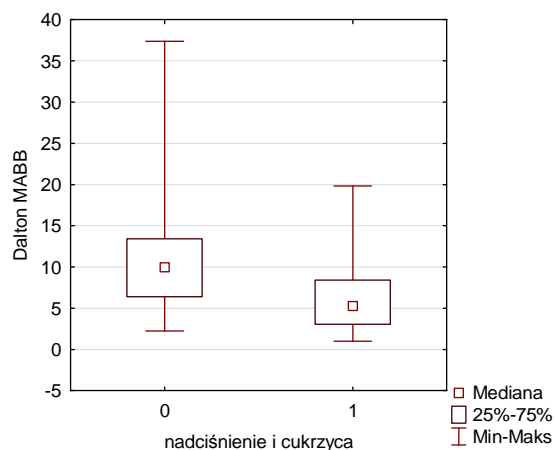
Wykres 53 rozkład wskaźnika zmienności krótkoterminowej van Geijna w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



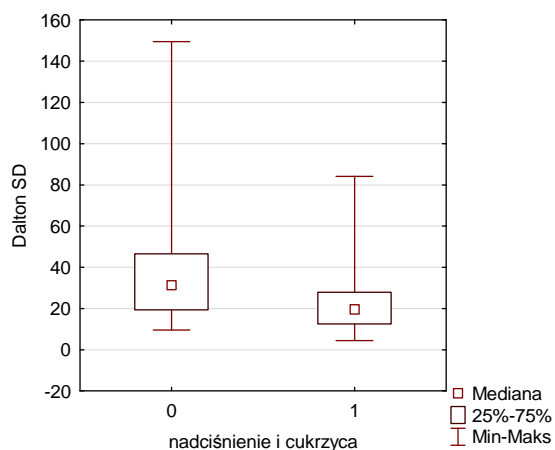
Wykres 54 rozkład wskaźnika zmienności krótkoterminowej Hueya w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



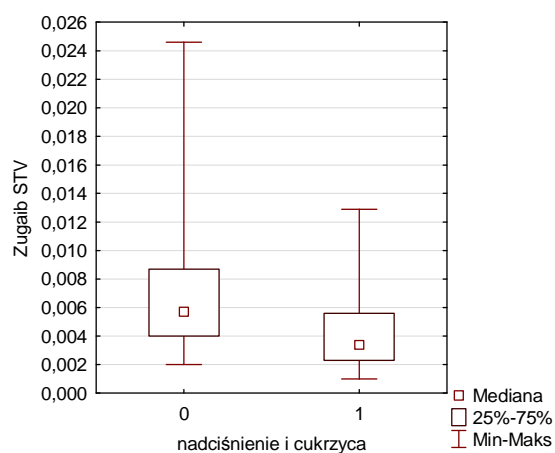
Wykres 55 rozkład wskaźnika zmienności długoterminowej Hueya w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



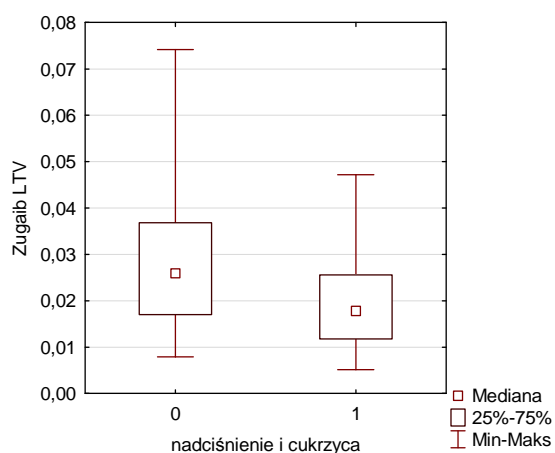
Wykres 56 rozkład wskaźnika zmienności krótkoterminowej Daltona w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



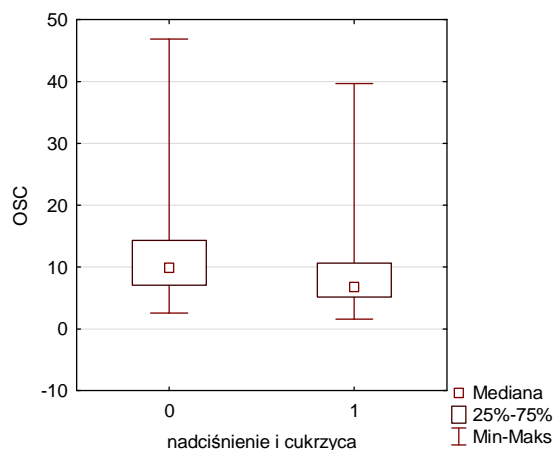
Wykres 57 rozkład wskaźnika zmienności długoterminowej Daltona w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



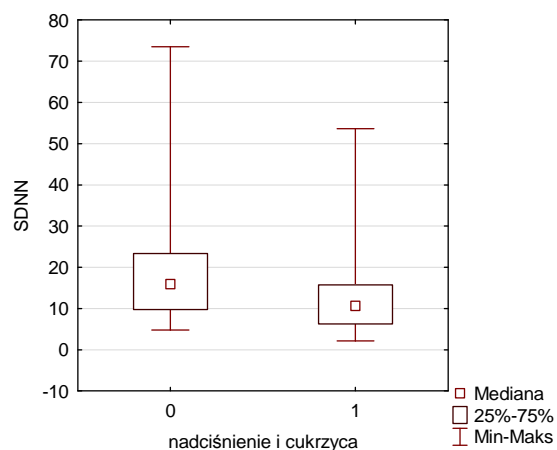
Wykres 58 rozkład wskaźnika zmienności krótkoterminowej Zugaiba w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



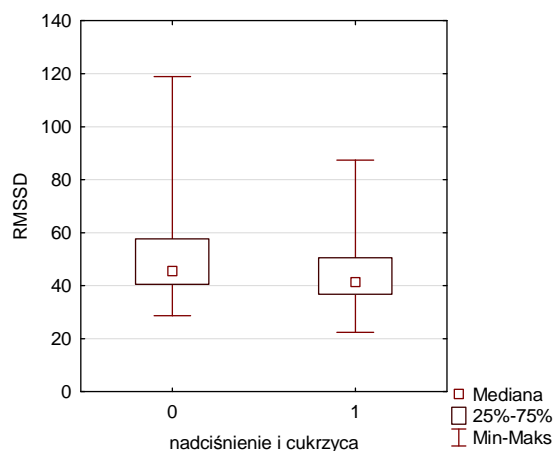
Wykres 59 rozkład wskaźnika zmienności długoterminowej Zugaiba w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



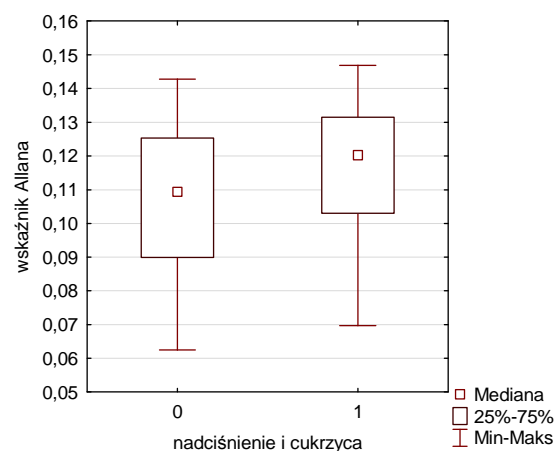
Wykres 60 rozkład wskaźnika szerokości wstęgi oscylacyjnej w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



Wykres 61 rozkład wskaźnika SDNN w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



Wykres 62 rozkład wskaźnika RMSSD w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)



Wykres 63 rozkład wskaźnika Allana w grupie osób chorujących na nadciśnienie tętnicze i cukrzycę i w grupie kontrolnej (0 – nie, 1 – tak)

a) kobiety

W celu sprawdzenia, czy istnieją statystycznie istotne różnice w wieku, BMI oraz wartościach wskaźników zmienności rytmu serca pomiędzy kobietami chorującymi na nadciśnienie tętnicze i cukrzycę oraz kobietami z grupy kontrolnej, z powodu braku zgodności z rozkładem normalnym, zastosowano test U Manna-Whitneya.

Nie wykazano istnienia istotnych różnic w wieku ($p=0,448$), w wartościach mierzonego w dziedzinie częstotliwości wskaźnika LF/HF ($p=0,705$). Kobiety chorujące na nadciśnienie tętnicze i cukrzycę nie różniły się od kobiet z grupy kontrolnej względem wskaźnika Allana ($p=0,088$) oraz wskaźnika RMSSD ($p=0,142$). Nie wykazano istnienia statystycznie istotnych różnic w wartościach, mierzonych w dziedzinie czasu, krótkoterminowych współczynników van Geijna i Hueya.

Kobiety chorujące na nadciśnienie tętnicze i cukrzycę mają istotnie wyższe wartości współczynnika BMI oraz istotnie niższe wartości wskaźników zmienności rytmu serca zmierzonych w dziedzinie czasu, zarówno krótko jak i długoterminowej. Szczegółowe dane dotyczące wartości poszczególnych parametrów oraz istotności statystycznej przedstawia **tabela 19**.

Tabela 19 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród kobiet chorujących na nadciśnienie tętnicze i cukrzycę oraz kobiet z grupy kontrolnej (test U Manna-Whitneya)

cecha	Kobiety bez zawału, nadciśnienia i cukrzycy (N=25)				Kobiety, które chorują na nadciśnienie i cukrzycę (N=19)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
BMI	25,0±3,2	25,2	20,0	31,6	29,7±4,2	29,2	19,7	35,9	0,001
de Haan STI	0,019±0,008	0,019	0,005	0,036	0,013±0,007	0,014	0,003	0,026	0,010
de Haan LTI	54,4±22,2	51,9	16,9	106,4	36,3±22,6	31,3	5,4	102,4	0,006
Yeh DI	0,016±0,009	0,013	0,004	0,046	0,009±0,005	0,010	0,002	0,019	0,006
Yeh II	0,038±0,017	0,034	0,010	0,076	0,024±0,011	0,022	0,007	0,048	0,007
Organ BAND	5,16±2,54	5,17	1,13	11,44	3,52±1,37	3,03	1,48	6,22	0,024
Huey LTV	96,76±45,44	94,17	27,23	185,86	71,02±35,18	62,55	21,51	139,45	0,049
Dalton MABB	11,01±6,37	10,26	2,25	36,43	6,37±3,92	6,38	1,03	13,79	0,004
Dalton SD	34,69±17,42	34,70	9,68	90,01	21,19±12,34	18,44	4,39	53,11	0,004
Zugaib STV	0,007±0,004	0,006	0,002	0,020	0,004±0,002	0,004	0,001	0,009	0,011
Zugaib LTV	0,029±0,012	0,027	0,008	0,050	0,019±0,009	0,018	0,005	0,038	0,009
OSC	12,3±6,9	11,6	2,5	33,0	8,0±3,0	7,3	3,2	14,0	0,020
SDNN	18,15±10,91	17,26	4,81	52,80	14,38±12,73	10,75	2,18	53,65	0,042

b) mężczyźni

W celu sprawdzenia, czy istnieją statystycznie istotne różnice w wieku, BMI oraz wartościach wskaźników zmienności rytmu serca pomiędzy mężczyznami chorującymi na nadciśnienie tętnicze i cukrzycę oraz mężczyznami z grupy kontrolnej, z powodu braku zgodności z rozkładem normalnym, zastosowano test U Manna-Whitneya.

Nie wykazano istnienia istotnych różnic w wartościach mierzonego w dziedzinie częstotliwości wskaźnika LF/HF ($p=0,750$). Mężczyźni chorujący na nadciśnienie tętnicze i cukrzycę nie różnili się od mężczyzn z grupy kontrolnej względem wskaźnika Allana ($p=0,087$) oraz wskaźnika RMSSD ($p=0,125$). Nie wykazano istnienia statystycznie istotnych różnic w wartościach, mierzonych w dziedzinie czasu, długoterminowych współczynników Organa ($p=0,075$) i Hueya ($p=0,094$) oraz krótkoterminowego współczynnika Hueya ($p=0,202$). Nie stwierdzono istnienia różnic w wartościach szerokości wstęgi oscylacyjnej ($p=0,117$).

Mężczyźni chorujący na nadciśnienie tętnicze i cukrzycę mają istotnie wyższe wartości współczynnika BMI oraz istotnie niższe wartości wskaźników zmienności rytmu serca zmierzonych w dziedzinie czasu, zarówno krótko jak i długoterminowej. Szczegółowe dane dotyczące wartości poszczególnych parametrów oraz istotności statystycznej przedstawia **tabela 20**.

Tabela 20 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród mężczyzn chorujących na nadciśnienie tętnicze i cukrzycę oraz mężczyzn z grupy kontrolnej (test U Manna-Whitneya)

cecha	Mężczyźni bez zawału, nadciśnienia i cukrzycy (N=37)				Mężczyźni, którzy chorują na nadciśnienie i cukrzycę (N=23)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
wiek	54,8±8,8	53,0	35,0	81,0	60,8±9,3	60,0	47,8	81,0	0,016
BMI	26,3±2,5	26,2	22,4	34,4	30,2±4,6	29,8	22,3	37,4	0,001
de Haan STI	0,018±0,009	0,015	0,005	0,039	0,012±0,010	0,009	0,004	0,050	0,010
de Haan LTI	55,2±31,9	46,0	16,5	159,4	36,0±20,0	34,2	10,48	97,4	0,009
Yeh DI	0,016±0,013	0,012	0,004	0,071	0,011±0,011	0,007	0,003	0,054	0,015
Yeh II	0,038±0,026	0,031	0,012	0,153	0,028±0,022	0,022	0,007	0,107	0,030
van Geijn ID	3,60±1,81	3,27	1,25	9,66	2,63±1,83	2,14	0,86	8,77	0,003
Dalton MABB	10,83±7,30	8,88	2,76	37,37	6,43±4,89	4,40	2,08	19,84	0,005
Dalton SD	36,10±25,71	28,05	10,67	149,49	23,57±17,91	20,66	7,63	84,16	0,011
Zugaib STV	0,007±0,005	0,005	0,002	0,025	0,004±0,003	0,003	0,001	0,013	0,009
Zugaib LTV	0,028±0,015	0,026	0,010	0,074	0,019±0,010	0,017	0,005	0,047	0,016
SDNN	19,31±15,55	13,91	5,35	73,55	13,33±10,88	10,54	3,82	42,63	0,031

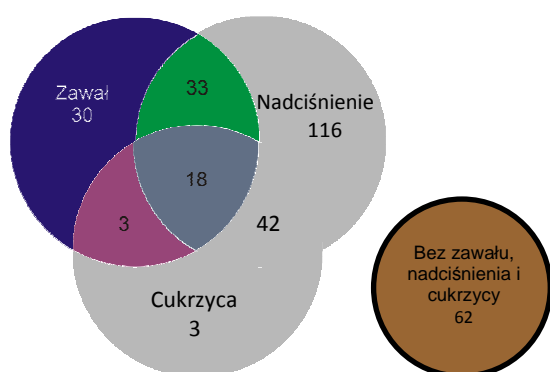
7.2 Regresja logistyczna

W modelu logistycznym, w celu odnalezienia minimum funkcji straty, zastosowano metodę quasi-Newtona jako metodę estymacji parametrów. Znany jest fakt, iż minimum funkcji należy spodziewać się w kierunku, w którym ona maleje. Metoda quasi-Newtona polega na oszacowaniu funkcji (w różnych punktach oraz w każdym kolejnym kroku) w celu estymacji pochodnych pierwszego i drugiego rzędu, które niosą za sobą informację o kierunku i szybkości zmian nachylenia funkcji.

W celu sprawdzenia istotności parametrów zastosowano test chi-kwadrat Walda. Aby ocenić, czy zbudowany model dobrze przewiduje zmienną zależną obliczono współczynnik R^2 Nagelkerke, który jest modyfikacją współczynnika pseudo R^2 .

Dla modeli przeprowadzono walidację, której celem było sprawdzenie jak model klasyfikuje nowe przypadki. W tym celu podzielono dane na zbiór uczący i testowy w stosunku 80% : 20%, czyli 80% przypadków z grupy badanej oraz 80% osób z grupy kontrolnej stanowiło zbiór uczący. Testowanie modelu odbyło się na grupie zbudowanej z 20% osób z grupy badanej i 20% osób z grupy kontrolnej. Podział na zbiór testowy i uczący był losowy, losowanie przypadków do zbioru uczącego i testowego odbyło się z wykorzystaniem generatora liczb losowych w programie Excel. W przypadku jednostek chorobowych, dla których zbudowano więcej niż jeden model istotny statystycznie, do walidacji wybrano model, który najlepiej klasyfikował pacjentów, czyli miał najwyższą czułość i najwyższą swoistość.

7.2.1 Pacjenci po przebytych zawałach mięśnia sercowego



Zbudowano cztery modele istotne statystycznie, które w mniejszym lub większym stopniu poprawnie klasyfikowały osoby po przebytych zawałach mięśnia sercowego i osoby z grupy kontrolnej (**tabela 21**). Poprawność klasyfikacji kształtowała się

na poziomie od 58,90% do 61,64%. Ilorazy szans wykonanych modeli przyjmowały wartości większe od 1 i wynosiły odpowiednio: 2,32 (dla modelu zbudowanego w oparciu o wskaźnik BMI oraz współczynnik zmienności krótkoterminowej de Haana), 2,65 (dla

modelu zbudowanego w oparciu o wskaźnik BMI oraz współczynnik zmienności długoterminowej de Haana), 1,78 (dla modelu zbudowanego w oparciu o wskaźnik BMI oraz współczynnik zmienności długoterminowej Yeha) oraz 2,17 (dla modelu zbudowanego w oparciu o wskaźnik BMI oraz współczynnik zmienności krótkoterminowej van Geijna), co wskazuje na to, że klasyfikacja przy użyciu modeli regresji logistycznej ma sens (**tabela 21**).

Wszystkie modele zbudowane zostały w oparciu o dwie zmienne. W każdym z nich istotnym parametrem okazał się współczynnik BMI. Drugą istotną zmienną był, mierzony w dziedzinie czasu, współczynnik de Haana (zarówno krótko jak i długoterminowy), wskaźnik zmienności długoterminowej Yeha i krótkoterminowej van Geijna (**tabela 21**). Jednostkowy iloraz szans dla współczynnika wagowo wzrostowego wynosi od 1,15 do 1,16 (w zależności od analizowanego modelu). 95 % przedziały ufności wyznaczone dla odpowiednich ilorazów szans nie zawierają 1, zatem można wysunąć przypuszczenie, że wraz ze wzrostem BMI rośnie szansa na wystąpienie zawału mięśnia sercowego. Dla istotnych statystycznie parametrów zmienności rytmu serca zmierzonych w dziedzinie czasu jednostkowe ilorazy szans kształtują się następująco:

- dla wskaźnika zmienności krótkoterminowej de Haana $OR=4,17*10^{-20}$, 95% przedział ufności: $(4,35*10^{-37}; 4,00*10^{-3})$;
- dla wskaźnika zmienności długoterminowej de Haana $OR=0,98$, 95% przedział ufności: $(0,97; 0,99)$;
- dla wskaźnika zmienności długoterminowej Yeha $OR=3,4*10^{-12}$, 95% przedział ufności: $(1,10*10^{-21}; 0,01)$;
- dla wskaźnika zmienności krótkoterminowej van Geijna $OR=0,71$, 95% przedział ufności: $(0,56; 0,89)$.

Ponieważ wszystkie ilorazy szans są mniejsze od 1 oraz odpowiednie przedziały ufności nie zawierają 1, więc można stwierdzić, że wraz ze wzrostem wartości tych wskaźników maleje szansa na wystąpienie zawału mięśnia sercowego. Szczegółowe dane dotyczące budowy omawianych modeli znajdują się w **Załączniku 2**.

Zbudowane modele dobrze klasyfikują osoby chore, gdyż czułość modeli mieściła się pomiędzy 73,81% a 78,57%, co wskazuje na to, iż osoby po zawale mięśnia sercowego co najmniej w 73,81% były właściwie zaklasyfikowane do grupy osób chorych. Swoistość modeli mieści się w przedziale od 38,71% do 41,92% (**tabela 21**).

Modele zbudowane przy użyciu metody regresji logistycznej są odpowiednie dla klasyfikacji osób z zawałem mięśnia sercowego, gdyż, z powodu poważnych następstw

zawału mających wpływ na dalsze życia pacjenta, wykrywalność osób chorych (wysoka czułość) jest istotniejsza od wysokiej swoistości.

Tabela 21 Modele regresji logistycznej klasyfikujące osoby po przeżytym zawale mięśnia sercowego i osoby z grupy kontrolnej

model	p	Iloraz szans modelu (OR)	Zmienne istotne	Iloraz szans dla parametrów	Klasyfikacja przypadków	R ² Nagelkerke
Zawał vs kontrolna (84 vs 62)	P=0,00079	2,32	BMI (p=0,006) de Haana STI (p=0,024)	1,15 <0,0001	%poprawnych:61,64 Czułość:78,57% Swoistość:38,71%	0,13
	*P=0,00046	2,65	BMI (p=0,003) de Haana LTI (p=0,014)	1,16 0,98	%poprawnych:63,01 Czułość:78,57% Swoistość:41,94%	0,13
	P=0,00041	1,78	BMI (p=0,004) Yeh II (p=0,017)	1,16 <0,0001	%poprawnych:58,90 Czułość:73,81% Swoistość:38,71%	0,14
	P=0,00008	2,17	BMI (p=0,004) van Geijn (p=0,003)	1,15 0,78	%poprawnych:60,96 Czułość:75,00% Swoistość:41,94%	0,16

*model wybrany do walidacji oraz dalszej analizy metodą drzew klasyfikacyjnych

Dla modelu zbudowanego w oparciu o zmienną BMI i wskaźnik zmienności długoterminowej de Haana, ponieważ cechował się on najlepszą klasyfikacją (najwyższy procent poprawnie zaklasyfikowanych przypadków, najwyższy OR oraz najlepsza czułość i swoistość), przeprowadzono walidację w celu sprawdzenia, jak model klasyfikuje nieznane, nowe przypadki. Zbiór uczący stanowiły 132 osoby, 56 osób z grupy kontrolnej i 76 osób po przebytym zawale mięśnia sercowego. Zbiór testowy liczył 14 osób: 6 osób z grupy kontrolnej i 8 osób po przebytym zawale mięśnia sercowego.

Dla zbioru uczącego uzyskano poprawność klasyfikacji na poziomie 61,36%. Czułość modelu wynosiła 75,00%, a swoistość 42,86%. Rozkład klasyfikacji przeprowadzonej na zbiorze uczącym przedstawia **tabela 22**. Dla zbioru testowego poprawność klasyfikacji wynosiła 71,43%. Uzyskano również wyższe wartości czułości (87,50%) oraz swoistości (50,00%) w porównaniu do modelu zbudowanego na zbiorze uczącym. Rozkład klasyfikacji przeprowadzonej na zbiorze testowym przedstawia **tabela 23**.

Tabela 22 Rozkład klasyfikacji na zbiorze uczącym przeprowadzonej przez model zbudowany w oparciu o zmienne BMI i de Haana LTI

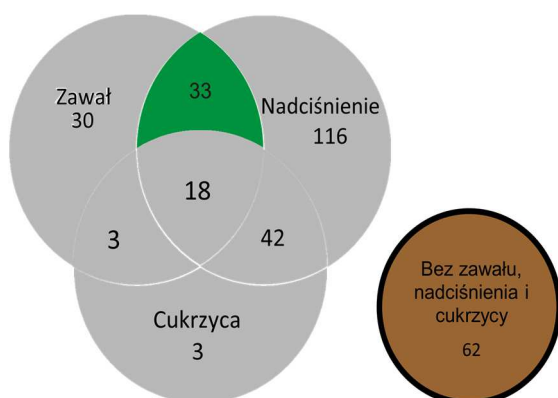
Zbiór uczący (N=132)				klasyfikacja
		przewidywane		
Obserwowane		0	1	%poprawnych: 61,36
	0	24	32	Swoistość: 42,86%
	1	19	57	Czułość: 75,00%

Tabela 23 Rozkład klasyfikacji na zbiorze testowym przeprowadzonej przez model zbudowany w oparciu o zmienne BMI i de Haana LTI

Zbiór testowy (N=14)				klasyfikacja
		przewidywane		
Obserwowane		0	1	%poprawnych: 71,43
	0	3	3	Swoistość: 50,00%
	1	1	7	Czułość: 87,50%

Wyniki klasyfikacji uzyskane na zbiorze testowym są lepsze niż na zbiorze uczącym: uzyskano wyższy procent poprawnie zaklasyfikowanych przypadków, wyższą swoistość oraz wyższą czułość, co pozwala wyciągnąć wniosek, że model będzie poprawnie klasyfikował nowe, nieznane przypadki.

7.2.2 Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie



Zbudowany model logistyczny w 78,95% poprawnie przyporządkowuje przypadki do klas. Iloraz szans tego modelu przyjmuje wartość większą od 1 i wynosi 12,67, co wskazuje na to, że klasyfikacja przy jego zastosowaniu ma sens (**tabela 24**).

Model został zbudowany w oparciu o siedem zmiennych. Parametrem istotnym statystycznie okazał się być wskaźnik wagowo – wzrostowy ($p=0,019$), dla którego jednostkowy iloraz szans wynosił 1,18. 95 % przedział ufności wyznaczony dla tego ilorazu nie zawierał 1, zatem można wysunąć przypuszczenie, że wraz ze wzrostem BMI rośnie szansa na wystąpienie zawału mięśnia sercowego i nadciśnienia tętniczego. Równie istotnym parametrem okazał się wiek pacjenta. Jednostkowy iloraz szans dla wieku wynosił 1,06, a wyznaczony 95% przedział ufności nie zawierał 1, zatem wraz ze wzrostem wieku rośnie ryzyko wystąpienia zawału mięśnia sercowego i nadciśnienia. Pozostałe parametry istotne statystycznie to: zmierzone w dziedzinie czasu krótkoterminowe wskaźniki zmienności rytmu serca de Haana, Yeha, van Geijna oraz długoterminowe Organa i Zugaiba (**tabela 24**). Dla istotnych statystycznie parametrów zmienności rytmu serca zmierzonych w dziedzinie czasu 95% przedziały ufności dla jednostkowych ilorazów szans kształtują się następująco:

- dla współczynnika zmienności krótkoterminowej de Haana ($1,37 \cdot 10^9$; ∞);
- dla współczynnika zmienności krótkoterminowej Yeha ($<10^{-8}$; $4,66 \cdot 10^{-5}$);
- dla współczynnika zmienności długoterminowej Organa (1,62; 10,03);
- dla współczynnika zmienności krótkoterminowej van Geijna (0,26; 0,99);
- dla współczynnika zmienności długoterminowej Zugaiba ($<10^{-37}$; $1,37 \cdot 10^{-34}$).

Ponieważ ilorazy szans dla wskaźnika de Haana oraz Organa są większe od 1, a odpowiednie przedziały ufności nie zawierają 1, można zatem stwierdzić, że wraz ze wzrostem wartości tych wskaźników rośnie szansa na wystąpienie zawału mięśnia sercowego przy współtowarzyszeniu nadciśnienia tętniczego. Dla wskaźnika zmienności krótkoterminowej Yeha, van Geijna oraz współczynnika zmienności długoterminowej

Zugałba ilorazy szans są mniejsze od 1. Przedziały ufności dla tych współczynników nie zawierają 1, więc wraz ze wzrostem wartości tych współczynników maleje szansa na wystąpienie zawału mięśnia sercowego przy współtowarzyszeniu nadciśnienia tętniczego. Szczegółowe dane dotyczące budowy omawianego modelu znajdują się w **Załączniku 2**.

Zaproponowany model dobrze klasyfikuje osoby z grupy kontrolnej, gdyż jego swoistość wynosiła 90,32%, co wskazuje na to, iż osoby z bez zawału, nadciśnienia i cukrzycy w ponad 90,00% były właściwie zaklasyfikowane do grupy kontrolnej. Niestety, model ma dość niską czułość, która wynosi 57,58% (**tabela 24**).

Model zbudowany przy użyciu metody regresji logistycznej jest odpowiedni dla klasyfikacji osób z zawałem mięśnia sercowego i nadciśnieniem tętniczym, gdyż, głównie z powodu nadciśnienia tętniczego, które jest chorobą przewlekłą i wymaga ciężkiego i długotrwałego leczenia, wykrywalność osób zdrowych (wysoka swoistość) jest bardzo istotna. Czuość modelu wynosząca nieco poniżej 60,00% nie jest zadowalająca, zwłaszcza, że z powodu poważnych następstw zawału mających wpływ na dalsze życia pacjenta, wykrywalność osób chorych (wysoka czułość) jest również istotna.

Tabela 24 Model regresji logistycznej klasyfikujący osoby po przebytym zawale mięśnia sercowego i chorujące na nadciśnienie tętnicze i osoby z grupy kontrolnej

model	p	Iloraz szans dla modelu (OR)	Zmienne istotne	Iloraz szans	Klasyfikacja przypadków	R ² Nagelkerke
Zawał i nadciśnienie vs kontrolna (33 vs 62)	P=0,00125	12,67	BMI (p=0,019) de Haana STI (p=0,025) Yeh DI (p=0,031) Organ (p=0,002) van Geijn ID (p=0,044) Zugaib LTV (p=0,003) Wiek (p=0,035)	1,18 <0,0001 4,03 0,51 <0,0001 1,06	%poprawnych:78,95 Czułość:57,58% Swoistość:90,32%	0,31

Dla powyższego modelu przeprowadzono walidację w celu sprawdzenia, jak model klasyfikuje nieznane, nowe przypadki. Zbiór uczący stanowiło 80 osób, 52 osób z grupy kontrolnej i 28 osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze. Zbiór testowy liczył 15 osób: 10 osób z grupy kontrolnej i 5 osób po przebytych zawale mięśnia sercowego i cierpiących na nadciśnienie tętnicze.

Dla zbioru uczącego uzyskano poprawność klasyfikacji na poziomie 80,00%. Czułość modelu wynosiła 64,29%, a swoistość 88,46%. Rozkład klasyfikacji przeprowadzonej na zbiorze uczącym przedstawia **tabela 25**. Dla zbioru testowego poprawność klasyfikacji wynosiła 73,33%. Uzyskano również nieco wyższe wartości swoistości (90,00%). Dość znacznie obniżyła się, w porównaniu do modelu zbudowanego na zbiorze uczącym, czułość modelu, która wynosiła 40,00%. Rozkład klasyfikacji przeprowadzonej na zbiorze testowym przedstawia **tabela 26**.

Tabela 25 Rozkład klasyfikacji na zbiorze uczącym przeprowadzonej przez model regresji logistycznej

Zbiór uczący (N=80)				klasyfikacja
		przewidywane		
Obserwowane		0	1	%poprawnych: 80,00
	0	46	6	Swoistość: 88,46%
	1	10	18	Czułość: 64,29%

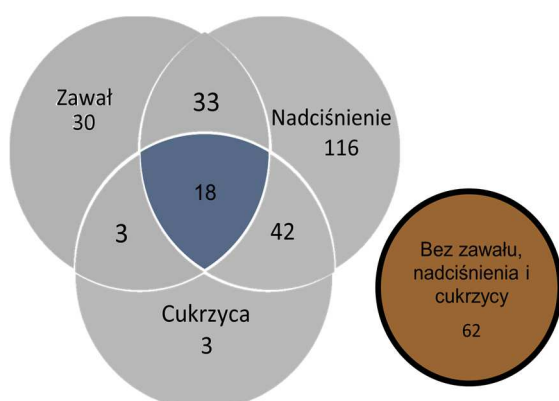
Tabela 26 Rozkład klasyfikacji na zbiorze testowym przeprowadzonej przez model regresji logistycznej

Zbiór testowy (N=15)				klasyfikacja
		przewidywane		
Obserwowane		0	1	%poprawnych: 73,33
	0	9	1	Swoistość: 90,00%
	1	3	2	Czułość: 40,00%

Wyniki klasyfikacji uzyskane na zbiorze testowym są zbliżone do wyników otrzymanych na zbiorze uczącym: uzyskano zbliżony procent poprawnie zaklasyfikowanych przypadków i nieco niższą swoistość, co pozwala wyciągnąć wniosek, że model będzie poprawnie klasyfikował nowe, nieznane przypadki. Czułość modelu opartego na zbiorze testowym jest

niższa o prawie 25 punktów procentowych od czułości modelu zbudowanego na zbiorze uczącym. Należy jednak zwrócić uwagę, że w grupie testowej było tylko 5 osób po przebytym zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze.

7.2.3 Pacjenci po przebytych zawałach mięśnia sercowego, którzy chorują na nadciśnienie i cukrzycę



Zbudowany model logistyczny w 85,00% poprawnie przyporządkowuje przypadki do klas. Iloraz szans tego modelu przyjmuje wartość większą od 1 i wynosi 24,00, co wskazuje na to, że klasyfikacja przy jego zastosowaniu ma sens (**tabela 27**).

Model został zbudowany w oparciu o dwie zmienne. Parametrem istotnym statystycznie okazał się być współczynnik BMI ($p=0,001$), dla którego jednostkowy iloraz szans wynosił 1,42. 95 % przedział ufności wyznaczony dla tego ilorazu nie zawierał 1, zatem można wysunąć przypuszczenie, że wraz ze wzrostem BMI rośnie szansa na wystąpienie zawału mięśnia sercowego przy współistnieniu nadciśnienia tętniczego i cukrzycy. Drugim parametrem, który uzyskał istotność statystyczną był współczynnik zmienności krótkoterminowej de Haana ($p=0,041$) (**tabela 27**). Jednostkowy iloraz szans dla wskaźnika de Haana STI wynosił $3,98 \cdot 10^{-38}$. Wyznaczony dla niego 95% przedział ufności wynosił ($<0,0001$; 0,03). Ponieważ nie zawiera on 1, więc wraz ze wzrostem wartości tego współczynnika maleje szansa na współwystąpienie analizowanych jednostek chorobowych. Szczegółowe dane dotyczące budowy omawianego modelu znajdują się w **Załączniku 2**.

Zaproponowany model dobrze klasyfikuje osoby z grupy kontrolnej, gdyż jego swoistość wynosiła 96,77%, co wskazuje na prawie bezbłędną klasyfikację osób bez zawału mięśnia sercowego, nadciśnienia tętniczego i cukrzycy do grupy kontrolnej. Czułość modelu wynosi 44,44% (**tabela 27**).

Zbudowany model może być z powodzeniem stosowany do klasyfikacji osób z zawałem mięśnia sercowego, nadciśnieniem tętniczym i cukrzycą, gdyż, zważywszy na przewlekłość nadciśnienia tętniczego i cukrzycy, które wymagają długotrwałego i intensywnego leczenia, wykrywalność osób zdrowych (wysoka swoistość) jest bardzo istotna. Czułość modelu wynosząca nieco ponad 44,00% nie jest zadowalająca, zwłaszcza, że z powodu poważnych następstw zawału mających wpływ na dalsze życie pacjenta, wykrywalność osób chorych (wysoka czułość) jest również istotna.

Tabela 27 Model regresji logistycznej klasyfikujący osoby po przebyłym zawale mięśnia sercowego i chorujące na nadciśnienie tętnicze i cukrzycę i osoby z grupy kontrolnej

model	p	Iloraz szans dla modelu (OR)	Zmienne istotne	Iloraz szans	Klasyfikacja przypadków	R² Nagelkerke
Zawał, nadciśnienie i cukrzyca vs kontrolna (18 vs 62)	P=0,0001	24	BMI (p=0,001) de Haana STI (p=0,038)	1,42 <0,0001	%poprawnych:85 Czułość:44,44% Swoistość:96,77%	0,37

Dla powyższego modelu przeprowadzono walidację w celu sprawdzenia, jak model klasyfikuje nieznane, nowe przypadki. Zbiór uczący stanowiło 58 osób, 45 osób z grupy kontrolnej i 13 osób po przebytym zawale mięśnia sercowego, chorujących na nadciśnienie tętnicze i cukrzycę. Zbiór testowy liczył 22 osoby: 17 osób z grupy kontrolnej i 5 osób po przebytym zawale mięśnia sercowego, cierpiących na nadciśnienie tętnicze i cukrzycę.

Dla zbioru uczącego uzyskano poprawność klasyfikacji na poziomie 81,03%. Czułość modelu wynosiła 53,85%, a swoistość 88,90%. Rozkład klasyfikacji przeprowadzonej na zbiorze uczącym przedstawia **tabela 28**. Dla zbioru testowego poprawność klasyfikacji wynosiła 72,73%. Uzyskano niższe wartości swoistości (76,47%). Na zbiorze testowym uzyskano czułość o 6 punktów procentowych wyższą niż na zbiorze uczącym. Rozkład klasyfikacji przeprowadzonej na zbiorze testowym przedstawia **tabela 29**.

Tabela 28 Rozkład klasyfikacji na zbiorze uczącym przeprowadzonej przez model regresji logistycznej

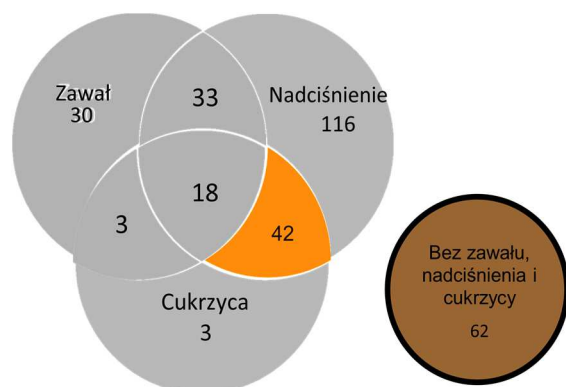
Zbiór uczący (N=58)				klasyfikacja
		przewidywane		
Obserwowane		0	1	%poprawnych: 81,03
	0	40	5	Swoistość: 88,90%
	1	6	7	Czułość: 53,85%

Tabela 29 Rozkład klasyfikacji na zbiorze testowym przeprowadzonej przez model regresji logistycznej

Zbiór testowy (N=22)				klasyfikacja
		przewidywane		
Obserwowane		0	1	%poprawnych: 72,73
	0	13	4	Swoistość: 76,47%
	1	2	3	Czułość: 60,00%

Wyniki klasyfikacji uzyskane na zbiorze testowym są zbliżone do wyników otrzymanych na zbiorze uczącym: uzyskano nieco niższy procent poprawnie zaklasyfikowanych i wyższą czułość, co pozwala wyciągnąć wniosek, że model będzie poprawnie klasyfikował nowe, nieznane przypadki. Swoistość, modelu opartego na zbiorze testowym jest niższa o prawie 12 punktów procentowych od swoistości modelu zbudowanego na zbiorze uczącym.

7.2.4 Pacjenci, którzy chorują na nadciśnienie i cukrzycę



Zbudowano cztery modele istotne statystycznie, które w mniejszym lub większym stopniu poprawnie klasyfikowały osoby chorujące na nadciśnienie tętnicze i cukrzycę oraz osoby z grupy kontrolnej (**tabela 30**). Poprawność klasyfikacji kształtowała się na poziomie od 70,19% do

82,69%. Ilorazy szans wykonanych modeli przyjmowały wartości większe od 1 i wynosiły odpowiednio: 22,14 (dla modelu zbudowanego w oparciu o wskaźnik BMI, wskaźniki zmienności krótkoterminowej de Haana, Yeha i van Geijna oraz współczynniki zmienności długoterminowej de Haana i Yeha), 9,57 (dla modelu zbudowanego w oparciu o wskaźnik BMI oraz współczynnik zmienności krótkoterminowej de Haana), 10,40 (dla modelu zbudowanego w oparciu o wskaźnik BMI oraz współczynnik zmienności długoterminowej Zugaiba) oraz 5,03 (dla modelu zbudowanego w współczynnik zmienności krótkoterminowej Daltona), co wskazuje na to, że klasyfikacja przy użyciu modeli regresji logistycznej ma sens (**tabela 30**).

W większości zaproponowanych modeli istotnym parametrem okazał się współczynnik BMI. Jednostkowy iloraz szans dla współczynnika wagowo wzrostowego wynosił od 1,37 do 1,43 (w zależności od analizowanego modelu). 95 % przedziały ufności wyznaczone dla odpowiednich ilorazów szans nie zawierają 1, zatem można wysunąć przypuszczenie, że wraz ze wzrostem BMI rośnie szansa na wystąpienie nadciśnienia tętniczego i cukrzycy.

W modelu zbudowanym w oparciu o największą liczbę zmiennych wskaźniki zmienności krótko i długoterminowej de Haana oraz współczynniki zmienności krótkoterminowej Yeha i van Geijna nie uzyskały istotności statystycznej. Pomimo tego są one niezbędne w modelu logistycznym i ich usunięcie powodowało uzyskiwanie braku istotności całego modelu. Dla istotnych statystycznie parametrów zmienności rytmu serca zmierzonych w dziedzinie czasu wchodzących w skład tego modelu jednostkowe 95% przedziały ufności dla ilorazów szans kształtują się następująco:

- dla wskaźnika zmienności długoterminowej Yeha (261270,10; ∞)
- dla wskaźnika zmienności długoterminowej Zugaiba ($<0,0001$; $4,51 \cdot 10^{-18}$).

Dla zmiennych wchodzących w skład pozostałych modeli 95% przedziały ufności dla ilorazów szans wynoszą odpowiednio:

- dla wskaźnika zmienności krótkoterminowej de Haana ($<0,0001$; $1,57 \cdot 10^{-5}$);
- dla wskaźnika zmienności długoterminowej Zugaiba ($<0,0001$; $2,98 \cdot 10^{-7}$);
- dla wskaźnika zmienności krótkoterminowej Daltona (0,76; 0,93).

Ponieważ wszystkie ilorazy szans są mniejsze od 1 oraz odpowiednie przedziały ufności nie zawierają 1, więc można stwierdzić, że wraz ze wzrostem wartości tych wskaźników maleje szansa na wystąpienie nadciśnienia tętniczego i cukrzycy. Szczegółowe dane dotyczące budowy omawianych modeli znajdują się w **Załączniku 2**.

Zbudowane modele dobrze klasyfikują osoby bez zawału mięśnia sercowego, nadciśnienia tętniczego i cukrzycy, gdyż swoistość modeli mieściła się pomiędzy 79,03% a 88,71%, co wskazuje na to, iż osoby z grupy kontrolnej prawie w 80,00% były właściwie zaklasyfikowane. Czulość modeli mieści się w przedziale od 57,14% do 73,81% (**tabela 30**).

Modele zbudowane przy użyciu metody regresji logistycznej są odpowiednie dla klasyfikacji osób z nadciśnieniem tętniczym i cukrzycą, gdyż, z powodu intensywnego i przewlekłego leczenia, wykrywalność osób nie chorujących na te schorzenia jest istotniejsza.

Tabela 30 Modele regresji logistycznej klasyfikujące osoby chorujące na nadciśnienie tętnicze i cukrzycę i osoby z grupy kontrolnej

model	p	Iloraz szans dla modelu (OR)	Zmienne istotne	Iloraz szans	Klasyfikacja przypadków	R ² Nagelkerke
Nadciśnienie i cukrzyca vs kontrolna (42 vs 62)	P<0,0001	22,14	BMI (p<0,0001) de Haana STI (p=0,343) de Haana LTI (p=0,212) Yeh DI (p=0,073) Yeh II (p=0,038) van Geijn (p=0,133) Zugaib LTV (p=0,031)	1,43 1,05 <0,0001 1,87 <0,0001	%poprawnych:82,69 Czułość:73,81% Swoistość:88,71%	0,49
	P<0,0001	9,57	BMI (p<0,0001) de Haana STI (p=0,018)	1,39 <0,0001	%poprawnych:75,96 Czułość:61,90% Swoistość:85,48%	0,40
	*P<0,0001	10,40	BMI (p<0,0001) Zugaib LTV (p=0,001)	1,37 <0,0001	%poprawnych:76,92 Czułość:66,67% Swoistość:83,87%	0,42
	P=0,00006	5,03	Dalton MABB (p=0,001)	0,84	%poprawnych:70,19 Czułość:57,14% Swoistość:79,03%	0,19

*model wybrany do walidacji oraz dalszej analizy metodą drzew klasyfikacyjnych

Dla modelu zbudowanego w oparciu o zmienną BMI i wskaźnik zmienności długoterminowej Zugaiba, ponieważ, cechował się on najlepszą klasyfikacją (najwyższy procent poprawnie zaklasyfikowanych przypadków, najwyższy OR oraz najlepsza czułość, pomijając model pierwszy), przeprowadzono walidację w celu sprawdzenia, jak model klasyfikuje nieznaną, nowe przypadki. Próba przeprowadzenia walidacji modelu zbudowanego w oparciu o największą liczbę zmiennych nie powiodła się. Model zbudowany na mniejszej liczbie przypadków nie uzyskał istotności statystycznej.

Zbiór uczący stanowiły 84 osoby, 50 osób z grupy kontrolnej i 34 osób chorujące na nadciśnienie tętnicze i cukrzycę. Zbiór testowy liczył 20 osób: 12 osób z grupy kontrolnej i 8 osób chorujących na nadciśnienie tętnicze i cukrzycę.

Dla zbioru uczącego uzyskano poprawność klasyfikacji na poziomie 71,38%. Czułość modelu wynosiła 61,76,00%, a swoistość 82,00%. Rozkład klasyfikacji przeprowadzonej na zbiorze uczącym przedstawia **tabela 31**. Dla zbioru testowego poprawność klasyfikacji wynosiła 90,00%. Uzyskano również wyższe wartości czułości (87,50%) oraz swoistości (91,67%) w porównaniu do modelu zbudowanego na zbiorze uczącym. Rozkład klasyfikacji przeprowadzonej na zbiorze testowym przedstawia **tabela 32**.

Tabela 31 Rozkład klasyfikacji na zbiorze uczącym przeprowadzonej przez wybrany model regresji logistycznej

Zbiór uczący (N=84)				klasyfikacja
		przewidywane		
Obserwowane		0	1	%poprawnych: 71,38
	0	41	9	Swoistość: 82,00%
	1	13	21	Czułość: 61,76%

Tabela 32 Rozkład klasyfikacji na zbiorze testowym przeprowadzonej przez wybrany model regresji logistycznej

Zbiór testowy (N=20)				klasyfikacja
		przewidywane		
Obserwowane		0	1	%poprawnych: 90,00
	0	11	1	Swoistość: 91,67%
	1	1	7	Czułość: 87,50%

Wyniki klasyfikacji uzyskane na zbiorze testowym są lepsze od wyników otrzymanych na zbiorze uczącym: uzyskano zdecydowanie wyższy procent poprawnie zaklasyfikowanych przypadków i wyższą swoistość, co pozwala wyciągnąć wniosek, że model będzie poprawnie klasyfikował nowe, nieznane przypadki. Czułość, modelu opartego na zbiorze testowym jest o ponad 25 punktów procentowych wyższa od czułości modelu zbudowanego na zbiorze uczącym.

7.2.5 Podsumowanie modeli regresji logistycznej

Tabela 33 podsumowanie modeli regresji logistycznej dla osób z przebyłym zawałem mięśnia sercowego, nadciśnieniem lub/i cukrzycą

	Osoby po przebyłym zawałe mięśnia sercowego	Osoby, po przebyłym zawałe mięśnia sercowego i chorujące na nadciśnienie tętnicze	Osoby, po przebyłym zawałe mięśnia sercowego i chorujące na nadciśnienie tętnicze i cukrzycę	Osoby, chorujące na nadciśnienie tętnicze i cukrzycę
Liczba zbudowanych modeli	4	1	1	4
Liczba osób z grupy kontrolnej	62	62	62	62
Liczba osób z grupy badanej	84	33	18	42
Procent błędnie zaklasyfikowanych przypadków dla wybranego modelu	36,99%	21,05%	15,00%	23,08%
Procent osób z grupy badanej błędnie zaklasyfikowanych	21,43%	42,42%	55,56%	33,33%
Procent osób z grupy kontrolnej błędnie zaklasyfikowanych	58,06%	9,68%	3,23%	16,13%
Iloraz szans dla wybranego modelu	2,65	8,35	24,00	10,40

7.3 Drzewa klasyfikacyjne

7.3.1 Pacjenci po przebytych zawale mięśnia sercowego

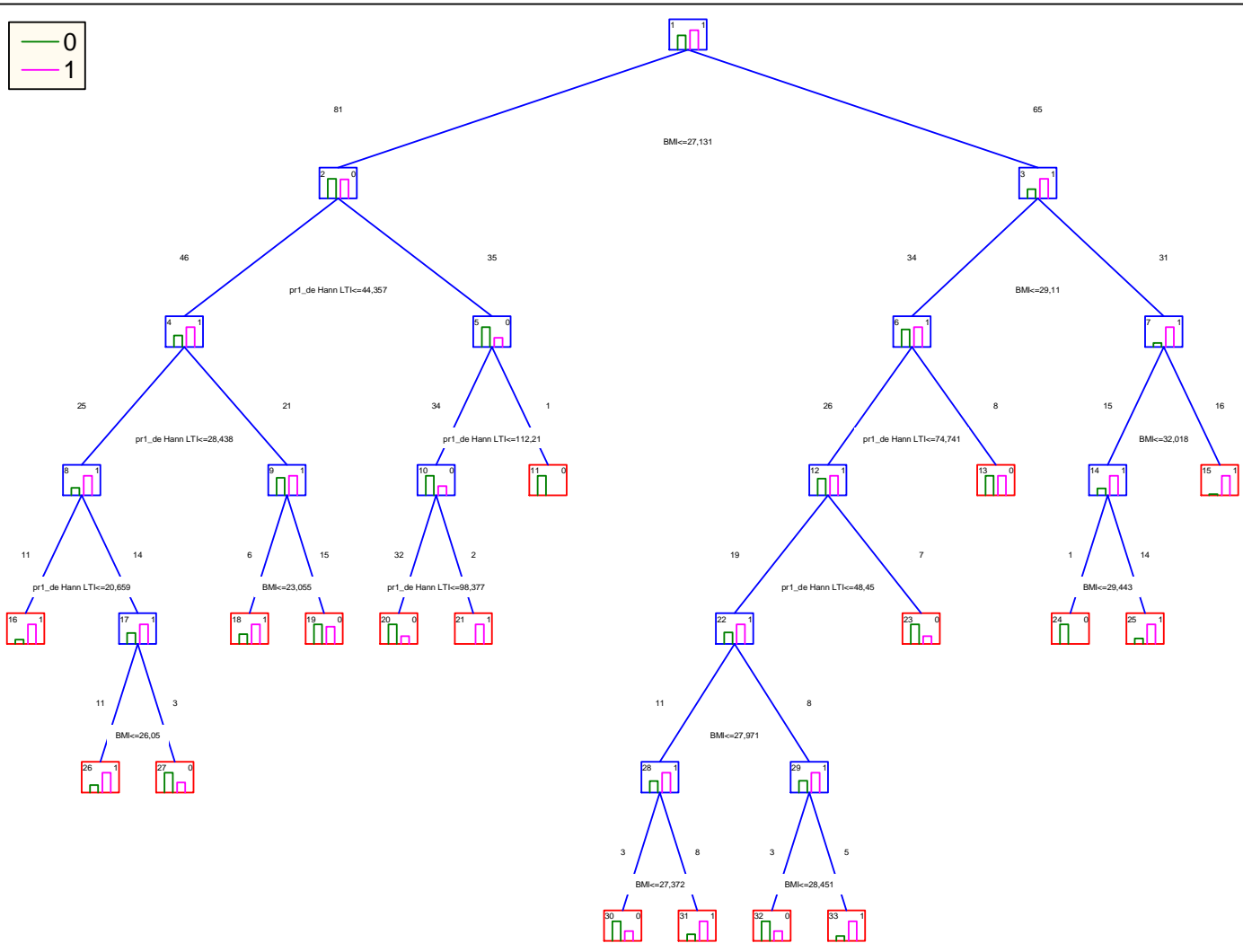
W celu wybrania podziałów w oparciu o zmienne predykcyjne: współczynnik wagowo - wzrostowy BMI i wskaźnik zmienności długoterminowej de Haana, zastosowano metodę dyskryminacyjnych podziałów jednowymiarowych, która jest dedykowana zarówno dla predyktorów nominalnych jak i porządkowych i polega na tym, że najpierw wyznacza się, który końcowy węzeł bieżącego drzewa najlepiej podzielić, a następnie dla wyznaczonego węzła znajduje się zmienną predykcyjną dla tego podziału. Jako miarę dobroci dopasowania wybrano miarę Giniego, przyjmującą wartość 0, gdy węzeł jest czysty, tzn. zawiera przypadki należące tylko do jednej kategorii zmiennej zależnej, a wartość maksymalną osiąga, gdy liczności klas w danym węźle są równe.

Jako regułę stopu wybrano regułę bezpośredniego zatrzymania typu FACT, która polega na dopuszczeniu do kontynuowania podziałów do czasu, gdy wszystkie węzły końcowe są czyste lub zawierają nie więcej przypadków niż określona minimalna frakcja liczności jednej lub więcej klas.

Zbudowane drzewo klasyfikacyjne w 67,12% poprawnie przewiduje przynależność przypadków do klas. Zmienna BMI stanowi pierwsze kryterium podziału. Jako stałą podziału przyjęto wartość 27,1. Kolejnym istotnym kryterium podziału dla osób przyporządkowanych do grupy kontrolnej jest współczynnik zmienności długoterminowej de Haana (stała podziału równa 44,357), a dla osób zaklasyfikowanych do grupy osób po przebytych zawale mięśnia sercowego ponownie współczynnik wagowo – wzrostowy, ze stałą podziału równą 29,1. Najistotniejszym predyktorem w tym drzewie był wskaźnik zmienności długoterminowej de Haana (**tabela 34**). Dokładną strukturę drzewa przedstawia **tabela A14 (Załącznik nr 3)** oraz **wykres 64**.

Globalne koszty sprawdzianu krzyżowego (średnie koszty dla prób testowych) są niższe od kosztów sprawdzianu krzyżowego dla wybranego drzewa. Podobnie, błąd standardowy globalnych kosztów sprawdzianu krzyżowego jest zbliżony do błędu standardowego kosztów wybranego drzewa (**tabela 35**). Świadczy to o tym, że wybrane drzewo jest właściwej wielkości i nie jest drzewem przeuczonym. Dodatkowo, koszty resubstytucji (koszty w próbie uczącej) są zbliżone do kosztów sprawdzianu krzyżowego dla wybranego drzewa, co potwierdza fakt, iż jest to drzewo najlepsze.

Czułość drzewa klasyfikacyjnego wynosi 70,24%, a swoistość 62,90%. Świadczy o tym, że model zbudowany w oparciu o metodę drzew klasyfikacyjnych może być z powodzeniem stosowany do wykrywania osób po przebytych zawale mięśnia sercowego, z powodu wysokiej czułości.



Wykres 64 Struktura drzewa klasyfikacyjnego – interpretacja graficzna (0 – nie, 1 – tak)

Tabela 34 Udział poszczególnych zmiennych w modelu CRT

0 – udział minimalny

100 – udział maksymalny

zmienna	ranking
BMI	93
de Haan LTI	100

Tabela 35 Błędy i macierze błędnych klasyfikacji w modelu CRT

liczba węzłów końcowych	koszty sprawdzianu krzyżowego	błąd std.	koszty resubstytucji
17	0,431507	0,048129	0,267123

Błędne klasyfikacje dla próby uczącej N=146		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	48	25
Przewidywane 1	14	59

Błędne klasyfikacje w globalnym sprawdzianie krzyżowym		
Globalne koszty=0,32877;		
odch. stand.=0,03888		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	39	25
Przewidywane 1	23	59

7.3.2 Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie

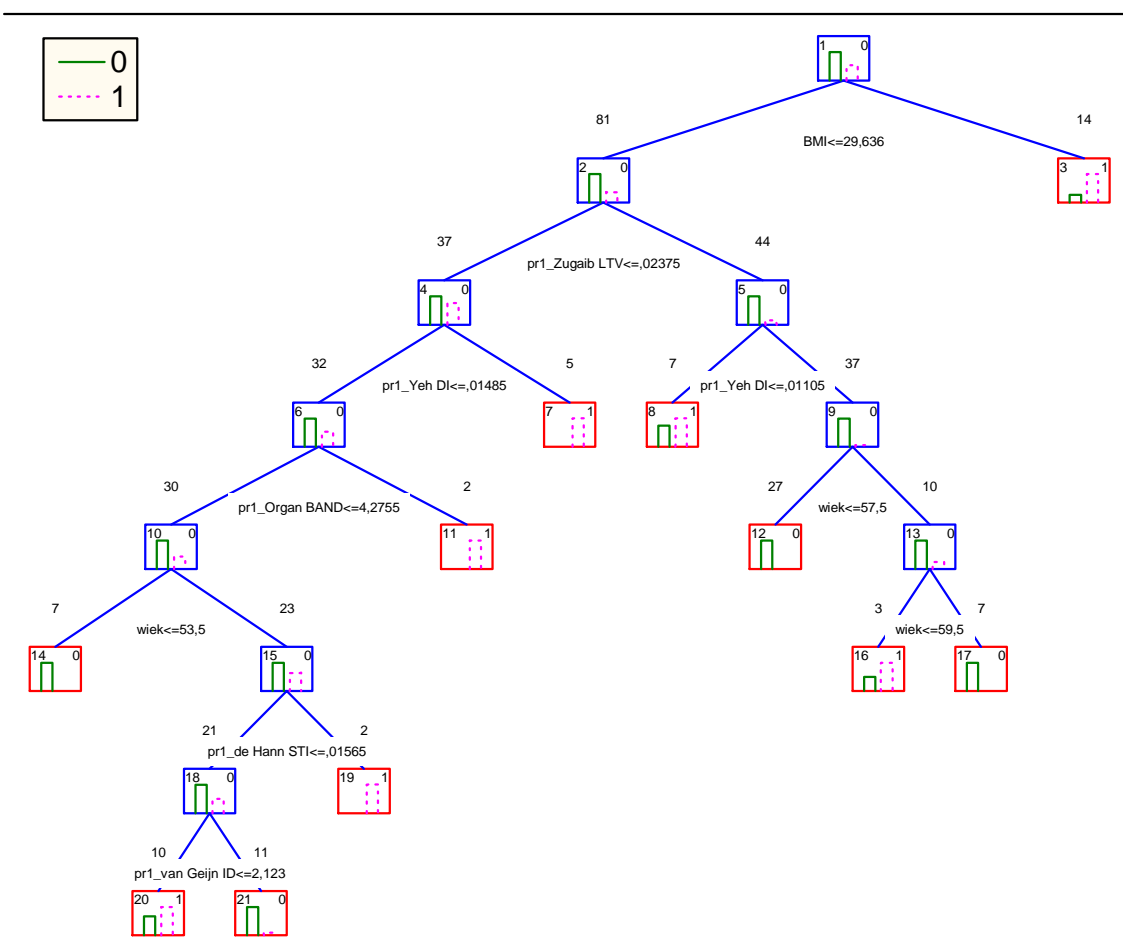
Drzewo klasyfikacyjne zbudowane w celu przyporządkowania osób po przebytych zawale mięśnia sercowego chorujących dodatkowo na nadciśnienie tętnicze i osób z grupy kontrolnej do właściwej klasy, oparto o zmienne wchodzące w skład modelu regresji logistycznej. Były to: współczynnik wagowo – wzrostowy, wiek, wskaźniki zmienności krótkoterminowej rytmu serca zmierzone w dziedzinie czasu de Haana, Yeha, van Geijna, współczynniki zmienności długoterminowej Organa i Zugaiba. W celu wybrania podziałów w oparciu o zmienne predykcyjne zastosowano metodę C&RT wyczerpującego poszukiwania podziałów jednowymiarowych, polegającą na sprawdzeniu wszystkich możliwych podziałów dla każdej zmiennej niezależnej tak, by znaleźć podział, przy którym następuje największa poprawa dobroci dopasowania.

Jako miarę dobroci dopasowania wybrano miarę Giniego, natomiast jako regułę stopu wybrano regułę bezpośredniego zatrzymania typu FACT, która polega na dopuszczeniu do kontynuowania podziałów do czasu, gdy wszystkie węzły końcowe są czyste lub zawierają nie więcej przypadków niż określona minimalna frakcja licznosci jednej lub więcej klas.

Zbudowane drzewo klasyfikacyjne w 64,21% poprawnie przewiduje przynależność przypadków do klas. Zmienna BMI stanowi pierwsze kryterium podziału. Jako stałą podziału przyjęto wartość 29,6. Kolejnym istotnym kryterium podziału był współczynnik zmienności długoterminowej Zugaiba (stała podziału równa 0,024), następnie wskaźnik zmienności krótkoterminowej Yeha, dalej wskaźnik zmienności długoterminowej Organa, wiek, współczynnik zmienności krótkoterminowej de Haana oraz van Geijna. Najistotniejszym predyktorem w tym drzewie był wskaźnik zmienności krótkoterminowej Yeha. Bardzo istotną zmienną był również współczynnik zmienności krótkoterminowej van Geijna. Najmniejszą wagę miał współczynnik wagowo – wzrostowy (**tabela 36**). Dokładną strukturę drzewa przedstawia **tabela A15 (Załącznik nr 3)** oraz **wykres 65**.

Globalne koszty sprawdzianu krzyżowego są niższe od kosztów sprawdzianu krzyżowego dla wybranego drzewa. Podobnie, błąd standardowy globalnych kosztów sprawdzianu krzyżowego jest niższy od błędu standardowego kosztów wybranego drzewa (**tabela 37**). Świadczy to o tym, że wybrane drzewo jest właściwej wielkości i nie jest drzewem przeuczonym. Dodatkowo, koszty resubstytucji (koszty w próbie uczącej) są zbliżone do kosztów sprawdzianu krzyżowego dla wybranego drzewa, co potwierdza fakt, iż jest to drzewo najlepsze.

Czułość drzewa klasyfikacyjnego wynosi 42,42%, a swoistość 75,81%. Świadczy o tym, że z uwagi na wysoką swoistość, model zbudowany w oparciu o metodę drzew klasyfikacyjnych może znaleźć zastosowanie w wykrywaniu osób, które nie przeszły zawału mięśnia sercowego, nie chorujących dodatkowo na nadciśnienie tętnicze.



Wykres 65 Struktura drzewa klasyfikacyjnego – interpretacja graficzna (0 – nie, 1 – tak)

Tabela 36 Udział poszczególnych zmiennych w modelu CRT

0 – udział minimalny

100 – udział maksymalny

zmienna	ranking
BMI	61
de Haan STI	84
Yeh DI	100
Organ Band	65
van Geijn ID	97
Zugaib LTV	86
wiek	90

Tabela 37 Błędy i macierze błędnych klasyfikacji w modelu CRT

liczba węzłów końcowych	koszty sprawdzianu krzyżowego	błąd std.	koszty resubstytucji
11	0,452632	0,060710	0,126316

Błędne klasyfikacje dla próby uczącej N=95		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	51	1
Przewidywane 1	11	32

Błędne klasyfikacje w globalnym sprawdzianie krzyżowym Globalne koszty=0,35789; odch. stand.=0,04918		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	47	19
Przewidywane 1	15	14

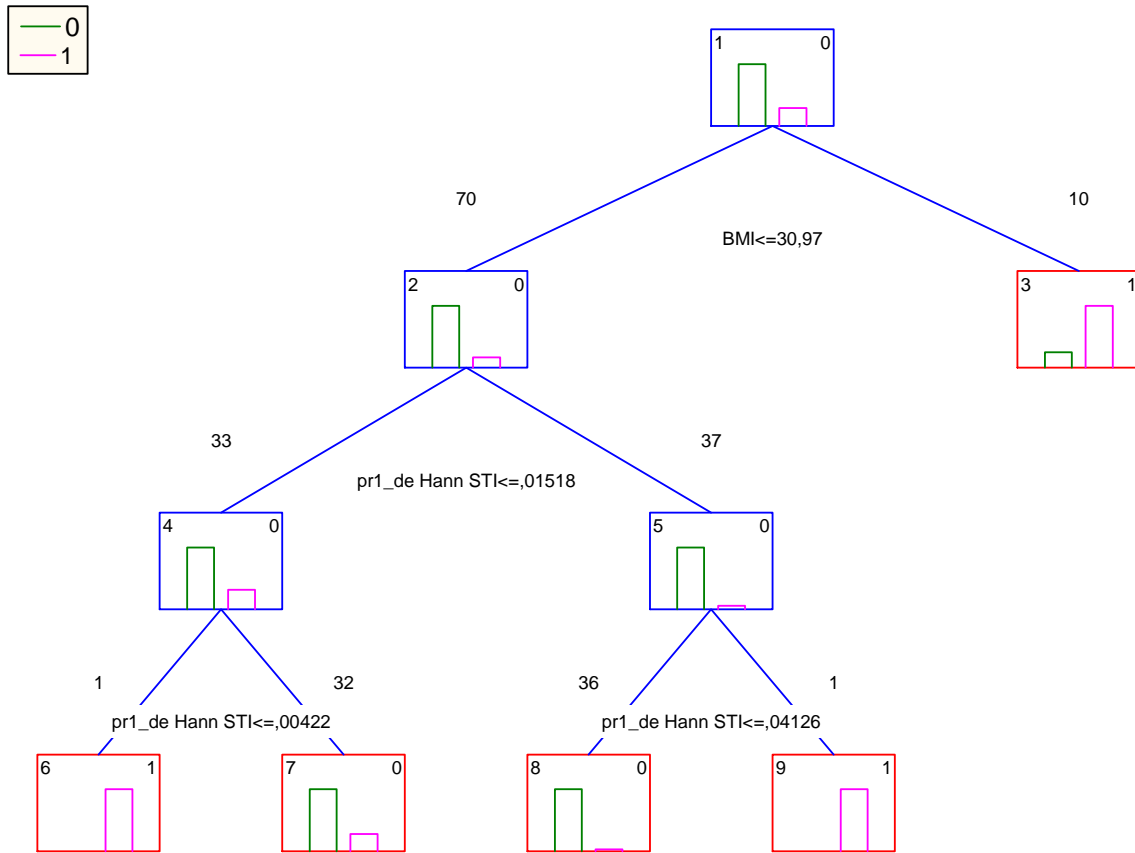
7.3.3 Pacjenci po przebytych zawale mięśnia sercowego, chorujący na nadciśnienie i cukrzycę

W celu przypisania osób po przebytych zawale mięśnia sercowego chorujących dodatkowo na nadciśnienie tętnicze i cukrzycę oraz osób z grupy kontrolnej do właściwej klasy, zbudowano drzewo klasyfikacyjne. Zbudowano je w oparciu o zmienne wchodzące w skład modelu regresji logistycznej. Były to: współczynnik wagowo – wzrostowy oraz wskaźnik zmienności krótkoterminowej de Haana. W celu wybrania podziałów w oparciu o zmienne predykcyjne zastosowano metodę dyskryminacyjnych podziałów jednowymiarowych, polegającą na tym, że wyznacza się który węzeł końcowy bieżącego drzewa najlepiej podzielić oraz zmiennej predykcyjnej dla tego podziału. Jako miarę dobroci dopasowania wybrano miarę Giniego, natomiast jako regułę stopu wybrano regułę bezpośredniego zatrzymania typu FACT, która dopuszcza kontynuowanie podziałów do czasu, gdy wszystkie węzły końcowe są czyste lub zawierają nie więcej przypadków niż określona minimalna frakcja licznosci jednej lub więcej klas.

Zbudowane drzewo klasyfikacyjne w 82,50% poprawnie przewiduje przynależność przypadków do klas. Zmienna BMI stanowi pierwsze kryterium podziału. Jako stałą podziału przyjęto wartość 30,9. Kolejnym istotnym kryterium podziału był współczynnik zmienności krótkoterminowej de Haana (stała podziału równa 0,015). Najistotniejszym predyktorem w tym drzewie był wskaźnik zmienności krótkoterminowej de Haana. W rankingu ważności predyktorów współczynnik wagowo - wzrostowy znalazł się na drugim miejscu (**tabela 38**). Dokładną strukturę drzewa przedstawia **tabela A16 (Załącznik nr 3)** oraz **wykres 66**.

Globalne koszty sprawdzianu krzyżowego są równe kosztom sprawdzianu krzyżowego dla wybranego drzewa. Natomiast błąd standardowy globalnych kosztów sprawdzianu krzyżowego jest niższy od błędu standardowego kosztów wybranego drzewa (**tabela 39**). Świadczy to o tym, że wybrane drzewo jest właściwej wielkości i nie jest drzewem przeuczonym. Dodatkowo, koszty resubstytucji (koszty w próbie uczącej) są zbliżone do kosztów sprawdzianu krzyżowego dla wybranego drzewa, co potwierdza fakt, iż jest to drzewo najlepsze.

Czułość drzewa klasyfikacyjnego wynosi 38,89%, a swoistość 95,16%. Świadczy o tym, że z uwagi na wysoką swoistość, model zbudowany w oparciu o metodę drzew klasyfikacyjnych może znaleźć zastosowanie w wykrywaniu osób po przebytych zawale mięśnia sercowego chorujących dodatkowo na nadciśnienie tętnicze i cukrzycę.



Wykres 66 Struktura drzewa klasyfikacyjnego – interpretacja graficzna (0 – nie, 1 – tak)

Tabela 38 Udział poszczególnych zmiennych w modelu CRT

0 – udział minimalny

100 – udział maksymalny

zmienna	ranking
BMI	72
de Haan STI	100

Tabela 39 Błędy i macierze błędnych klasyfikacji w modelu CRT

liczba węzłów końcowych	koszty sprawdzianu krzyżowego	błąd std.	koszty resubstytucji
5	0,175000	0,044282	0,125000

Błędne klasyfikacje dla próby uczącej N=80		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	60	8
Przewidywane 1	2	10

Błędne klasyfikacje w globalnym sprawdzianie krzyżowym		
Globalne koszty=0,175; odch. stand.=0,04248		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	59	11
Przewidywane 1	3	7

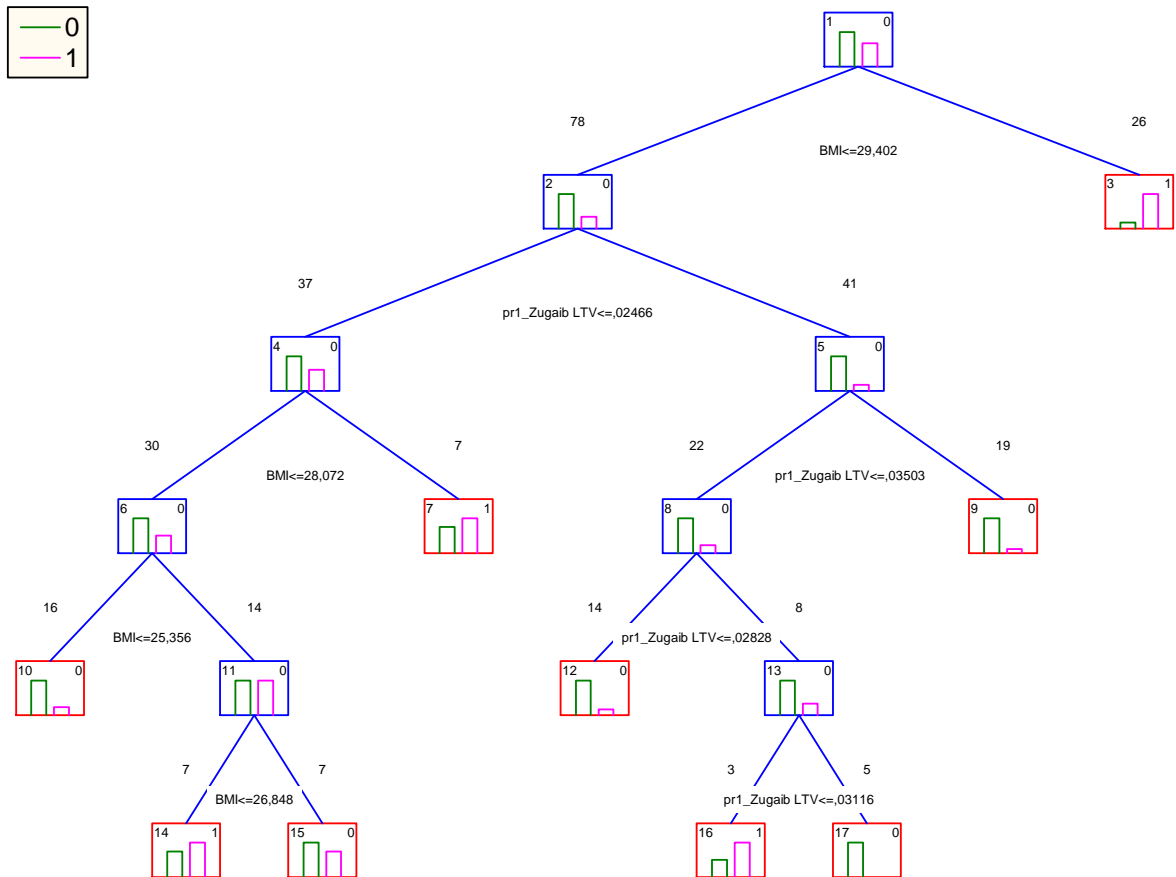
7.3.4 Pacjenci chorujący na nadciśnienie i cukrzycę

Drzewo klasyfikacyjne, zbudowane w celu klasyfikacji osób chorujących na nadciśnienie tętnicze i cukrzycę oraz osób z grupy kontrolnej do właściwej klasy, zbudowano w oparciu o współczynnik wagowo – wzrostowy oraz wskaźnik zmienności długoterminowej Zugaiba. W celu wybrania podziałów w oparciu o zmienne predykcyjne zastosowano metodę dyskryminacyjnych podziałów jednowymiarowych, w której wyznaczono, który węzeł końcowy bieżącego drzewa najlepiej podzielić oraz wybrano zmienną predykcyjną dla tego podziału. Jako miarę dobroci dopasowania wybrano miarę Giniego, natomiast jako regułę stopu wybrano regułę bezpośredniego zatrzymania typu FACT, która dopuszcza kontynuowanie podziałów do czasu, gdy wszystkie węzły końcowe są czyste lub zawierają nie więcej przypadków niż określona minimalna frakcja liczości jednej lub więcej klas.

Zbudowane drzewo klasyfikacyjne w 72,12% poprawnie przewiduje przynależność przypadków do klas. Zmienna BMI stanowi pierwsze kryterium podziału. Jako stałą podziału przyjęto wartość 29,4. Kolejnym istotnym kryterium podziału był współczynnik zmienności długoterminowej Zugaiba (stała podziału równa 0,025). Najistotniejszym predyktorem w tym drzewie był współczynnik BMI. Wskaźnik zmienności długoterminowej Zugaiba znalazł się na drugim miejscu (**tabela 40**). Dokładną strukturę drzewa przedstawia **tabela A17 (Załącznik nr 3)** oraz **wykres 67**.

Globalne koszty sprawdzianu krzyżowego są równe kosztom sprawdzianu krzyżowego dla wybranego drzewa. Natomiast błąd standardowy globalnych kosztów sprawdzianu krzyżowego jest niższy od błędu standardowego kosztów wybranego drzewa (**tabela 41**). Świadczy to o tym, że wybrane drzewo jest właściwej wielkości i nie jest drzewem przeuczonym. Dodatkowo, koszty resubstytucji (koszty w próbie uczącej) są zbliżone do kosztów sprawdzianu krzyżowego dla wybranego drzewa, co potwierdza fakt, iż jest to drzewo najlepsze.

Czułość drzewa klasyfikacyjnego wynosi 73,81%, a swoistość 70,97%. Świadczy o tym, że z uwagi na wysoką swoistość, Model zbudowany w oparciu o metodę drzew klasyfikacyjnych może znaleźć zastosowanie w wykrywaniu osób chorujących na nadciśnienie tętnicze i cukrzycę, gdyż cechuje się wysoką czułością i swoistością.



Wykres 67 Struktura drzewa klasyfikacyjnego – interpretacja graficzna (0 – nie, 1– tak)

Tabela 40 Udział poszczególnych zmiennych w modelu CRT

0 – udział minimalny

100 – udział maksymalny

zmienna	ranking
BMI	100
de Haan STI	94

Tabela 41 Błędy i macierze błędnych klasyfikacji w modelu CRT

liczba węzłów końcowych	koszty sprawdzianu krzyżowego	błąd std.	koszty resubstytucji
9	0,27885	0,04803	0,201923

Błędne klasyfikacje dla próby uczącej N=104		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	51	10
Przewidywane 1	11	32

Błędne klasyfikacje w globalnym sprawdzianie krzyżowym		
Globalne koszty=0,27885; odch. stand.=0,04397		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	44	11
Przewidywane 1	18	31

7.3.5 Podsumowanie modeli drzew klasyfikacyjnych

Tabela 42 podsumowanie modeli drzew klasyfikacyjnych dla osób z przebyłym zawałem mięśnia sercowego, nadciśnieniem lub/i cukrzycą

	Osoby po przebyłym zawałe mięśnia sercowego	Osoby, po przebyłym zawałe mięśnia sercowego i chorujące na nadciśnienie tętnicze	Osoby, po przebyłym zawałe mięśnia sercowego i chorujące na nadciśnienie tętnicze i cukrzycę	Osoby, chorujące na nadciśnienie tętnicze i cukrzycę
Zmienne w modelu	BMI de Haan LTI	BMI de Haan STI Yeh DI Organ BAND van Geijn ID Zugaib LTV wiek	BMI de Haan STI	BMI Zugaib LTV
Liczba węzłów końcowych	17	11	5	9
Koszt globalnego sprawdzianu krzyżowego	32,88%	35,79%	17,50%	27,88%
Liczba osób z grupy kontrolnej	62	62	62	62
Liczba osób z grupy badanej	84	33	18	42
Procent osób z grupy badanej błędnie zaklasyfikowanych	29,76%	57,58%	61,11%	26,19%
Procent osób z grupy kontrolnej błędnie zaklasyfikowanych	37,10%	24,19%	4,84%	29,03%

7.4 Podsumowanie modeli

Tabela 43 Podsumowanie dwóch wybranych modeli analizowanych w pracy dla osób po przebytych zawale mięśnia sercowego oraz dla osób po przebytych zawale mięśni sercowego i dodatkowo chorujących na nadciśnienie

klasyfikacja	Zawał vs kontrolna		Zawał i nadciśnienie vs kontrolna	
	Regresja logistyczna	Drzewa klasyfikacyjne	Regresja logistyczna	Drzewa klasyfikacyjne
% poprawnie sklasyfikowanych	63,01	67,12	78,95	64,21
czułość (%)	78,57	70,24	57,58	42,42
swoistość (%)	41,94	62,90	90,32	75,81

Tabela 44 Podsumowanie dwóch wybranych modeli analizowanych w pracy dla osób po przebytych zawale mięśni sercowego, dodatkowo chorujących na nadciśnienie i cukrzycę oraz osób chorujących na nadciśnienie i cukrzycę

klasyfikacja	Zawał, nadciśnienie i cukrzyca vs kontrolna		Nadciśnienie i cukrzyca vs kontrolna	
	Regresja logistyczna	Drzewa klasyfikacyjne	Regresja logistyczna	Drzewa klasyfikacyjne
% poprawnie sklasyfikowanych	85,00	82,50	76,92	72,12
czułość (%)	44,44	38,89	66,67	73,81
swoistość (%)	96,77	95,16	83,87	70,97

Poprawna klasyfikacja przypadków dla regresji logistycznej mieściła się w przedziale od 61,01% do 85,00%, w zależności od analizowanej grupy badanej. Czuość zbudowanych modeli kształtowała się na poziomie od 44,44% do 78,57%, a swoistość od 41,94% do 96,77%. Modele zbudowane z wykorzystaniem drzew klasyfikacyjnych cechowały się poprawnością klasyfikacji na poziomie od 64,21% do 82,50%. Identyfikacja osób z grupy badanej mieściła się w przedziale od 38,89% do 73,81%, w zależności od analizowanego schorzenia. Swoistość zbudowanych modeli wynosiła od 59,68% do 95,16% (tabela 43 i tabela 44).

W celu oceny skuteczności omawianych modeli zastosowano dodatkowo metodę DATA MINING – szybkie wdrażanie modeli predykcyjnych. Wymaga ona, aby w obydwu stosowanych technikach użyte były te same zmienne predykcyjne. Jest to zgodne z ideą przyjętą w pracy, która zakładała, że aby móc porównać dwie różne metody, należy do tego celu zastosować te same układy predyktorów.

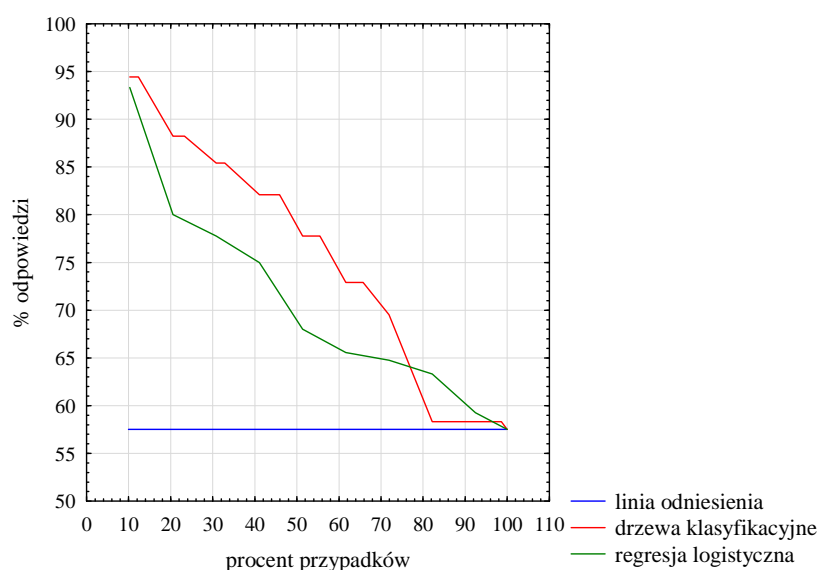
7.4.1 Osoby po przebytych zawale mięśnia sercowego

Zmienne niezależne użyte do budowy modeli klasyfikujących osoby po przebytych zawale mięśnia sercowego oraz osoby z grupy kontrolnej to BMI oraz wskaźnik zmienności długoterminowej de Haana.

Tabela 45 Procent błędnych klasyfikacji

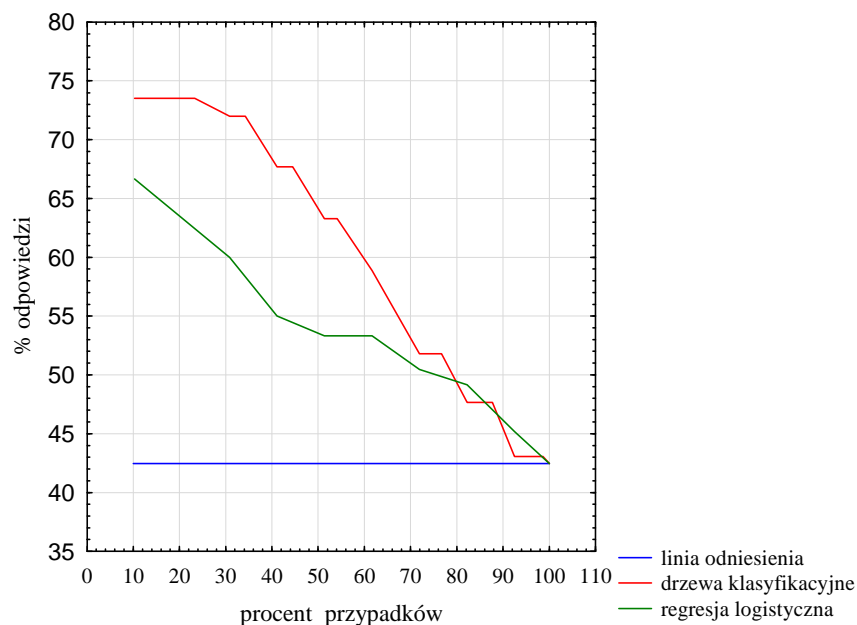
	drzewa klasyfikacyjne	regresja logistyczna
błąd	26,71%	36,99%

Regresja logistyczna cechuje się wyższym błędem klasyfikacji w porównaniu do modelu drzew klasyfikacyjnych. Procent błędnych klasyfikacji w modelu regresji jest o ponad 10 punktów procentowych wyższy w porównaniu do modelu drzew klasyfikacyjnych (**tabela 45**).



Wykres 68 Procent prawidłowych odpowiedzi dla osób po przebytych zawale mięśnia sercowego

Z wykresu 68 można odczytać, iż biorąc 10 procent przypadków najpewniej zaklasyfikowanych przez drzewa klasyfikacyjne do grupy osób po przebytych zawale mięśnia sercowego otrzymamy próbkę, w której co najmniej 94 procent przypadków należy do wybranej klasy (do grupy osób po przebytych zawale). Dla modelu regresji logistycznej można odczytać, że biorąc 10 procent przypadków najpewniej zaklasyfikowanych do grupy osób po przebytych zawale mięśnia sercowego otrzymuje się próbkę, w której 93 procent przypadków należy do grupy osób po przebytych zawale mięśnia sercowego. Linia odniesienia na tym wykresie jest pozioma i odpowiada klasyfikacji losowej; można jej użyć do porównania przydatności modeli klasyfikacyjnych.



Wykres 69 Procent prawidłowych odpowiedzi dla osób z grupy kontrolnej

Z wykresu 69 można odczytać, iż biorąc 20 procent przypadków najpewniej zaklasyfikowanych przez drzewa klasyfikacyjne do grupy osób bez zawału mięśnia sercowego, nadciśnienia i cukrzycy otrzymamy próbkę, w której co najmniej 73 procent przypadków należy do wybranej klasy (do grupy kontrolnej). Dla modelu regresji logistycznej można odczytać, że biorąc 10 procent przypadków najpewniej zaklasyfikowanych do grupy osób bez zawału mięśnia sercowego, nadciśnienia i cukrzycy otrzymuje się próbkę, w której 67 procent przypadków należy do grupy kontrolnej. Linia odniesienia na tym wykresie odpowiada klasyfikacji losowej.

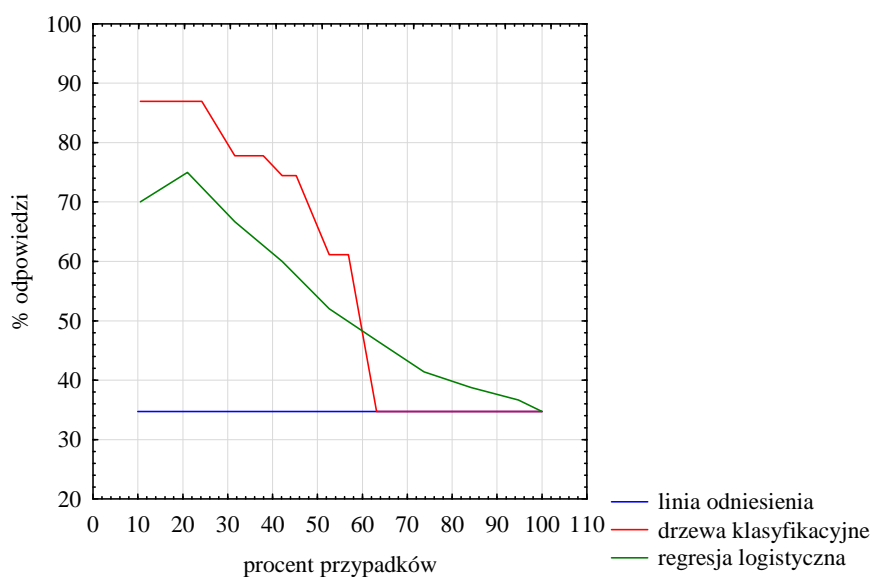
7.4.2 Osoby po przebytych zawałach mięśnia sercowego i chorujące na nadciśnienie

Zmienne niezależne użyte do budowy modeli klasyfikujących osoby po przebytych zawałach mięśnia sercowego i chorujące na nadciśnienie oraz osoby z grupy kontrolnej to BMI, wiek oraz wskaźniki zmienności krótkoterminowej: de Haana, Yeha, van Geijna, wskaźniki zmienności długoterminowej: Organa i Zugaiba.

Tabela 46 Procent błędnych klasyfikacji

	drzewa klasyfikacyjne	regresja logistyczna
błąd	12,63%	21,05%

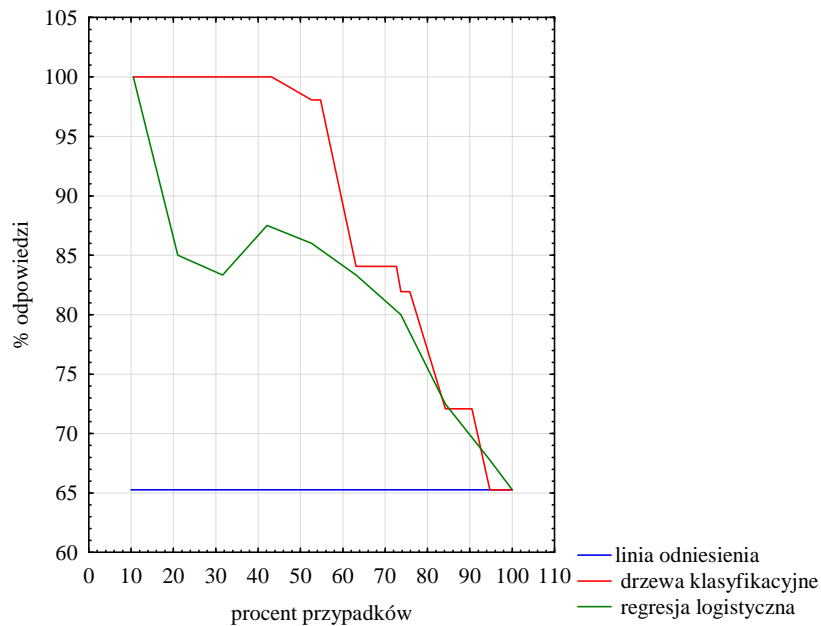
Regresja logistyczna cechuje się wyższym błędem klasyfikacji w porównaniu do modelu drzew klasyfikacyjnych. Procent błędnych klasyfikacji w modelu regresji jest o prawie 10 punktów procentowych wyższy w porównaniu do modelu drzew klasyfikacyjnych (**tabela 46**).



Wykres 70 Procent prawidłowych odpowiedzi dla osób po przebytych zawałach mięśnia sercowego i chorujących na nadciśnienie

Z wykresu 70 można odczytać, iż biorąc 10 procent przypadków najpewniej zaklasyfikowanych przez drzewa klasyfikacyjne do grupy osób po przebytych zawałach mięśnia sercowego i chorujących na nadciśnienie otrzymamy próbkę, w której 88 procent przypadków należy do wybranej klasy (do grupy osób po przebytych zawałach i z

nadciśnieniem tętniczym). Dla modelu regresji logistycznej można odczytać, że biorąc 10 procent przypadków najpewniej zaklasyfikowanych do grupy osób po przebytych zawale mięśnia sercowego otrzymuje się próbkę, w której 70 procent przypadków należy do grupy osób po przebytych zawale mięśnia sercowego i cierpiących na nadciśnienie. Linia odniesienia na tym wykresie odpowiada klasyfikacji losowej.



Wykres 71 Procent prawidłowych odpowiedzi dla osób z grupy kontrolnej

Z wykresu 71 można odczytać, iż biorąc 10 procent przypadków najpewniej zaklasyfikowanych przez regresję logistyczną do grupy osób bez zawału mięśnia sercowego, nadciśnienia i cukrzycy otrzymamy próbkę, w której 100 procent przypadków należy do wybranej klasy (do grupy kontrolnej). Dla modelu drzew klasyfikacyjnych można odczytać, że biorąc 40 procent przypadków najpewniej zaklasyfikowanych do grupy osób bez zawału mięśnia sercowego, nadciśnienia i cukrzycy otrzymuje się próbkę, w której 100 procent przypadków należy do grupy kontrolnej. Linia odniesienia na tym wykresie odpowiada klasyfikacji losowej.

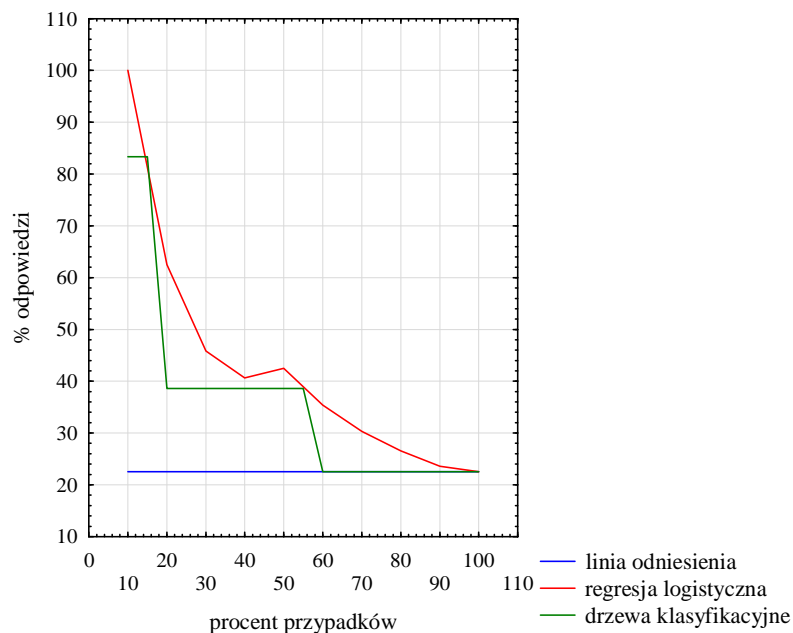
7.4.3 Osoby po przebytych zawałach mięśnia sercowego oraz chorujące na nadciśnienie i cukrzycę

Zmienne niezależne użyte do budowy modeli klasyfikujących osoby po przebytych zawałach mięśnia sercowego i chorujące na nadciśnienie oraz cukrzycę i osoby z grupy kontrolnej to BMI oraz wskaźnik zmienności krótkoterminowej de Haana.

Tabela 47 Procent błędnych klasyfikacji

	drzewa klasyfikacyjne	regresja logistyczna
błąd	12,50%	15,00%

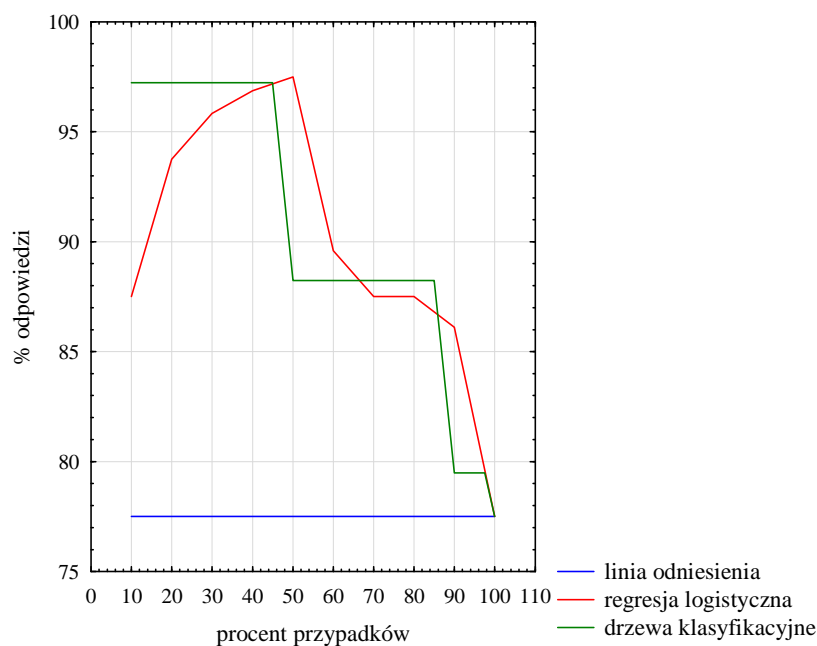
Regresja logistyczna cechuje się nieznacznie wyższym błędem klasyfikacji w porównaniu do modelu drzew klasyfikacyjnych. Procent błędnych klasyfikacji w modelu regresji jest o 2,5 punktu procentowego wyższy w porównaniu do modelu drzew klasyfikacyjnych i wynosi 15,00 procent (**tabela 47**).



Wykres 72 Procent prawidłowych odpowiedzi dla osób po przebytych zawałach mięśnia sercowego i chorujących na nadciśnienie tętnicze oraz cukrzycę

Z wykresu 72 można odczytać, iż biorąc 10 procent przypadków najpewniej zaklasyfikowanych przez regresję logistyczną do grupy osób po przebytych zawałach mięśnia

sercowego i chorujących na nadciśnienie oraz cukrzycę otrzymamy próbkę, w której 100 procent przypadków należy do wybranej klasy (do grupy osób po przebytym zawale i z nadciśnieniem tętniczym oraz cukrzycą). Dla modelu drzew klasyfikacyjnych można odczytać, że biorąc 10 procent przypadków najpewniej zaklasyfikowanych do grupy osób po przebytym zawale mięśnia sercowego otrzymuje się próbkę, w której 84 procent przypadków należy do grupy osób po przebytym zawale mięśnia sercowego i cierpiących na nadciśnienie oraz cukrzycę. Linia odniesienia na tym wykresie odpowiada klasyfikacji losowej.



Wykres 73 Procent prawidłowych odpowiedzi dla osób z grupy kontrolnej

W przypadku osób z grupy kontrolnej, z wykresu 73 można odczytać, że biorąc 40% przypadków najpewniej przyporządkowanych przez drzewa klasyfikacyjne do grupy osób bez zawału mięśnia sercowego, nadciśnienia i cukrzycy otrzymujemy próbkę, w której 97 procent przypadków należy do grupy kontrolnej. Natomiast w przypadku modelu regresji logistycznej wynik taki uzyskuje się, biorąc 50 procent przypadków najpewniej zaklasyfikowanych do grupy kontrolnej. Linia odniesienia odpowiada klasyfikacji losowej.

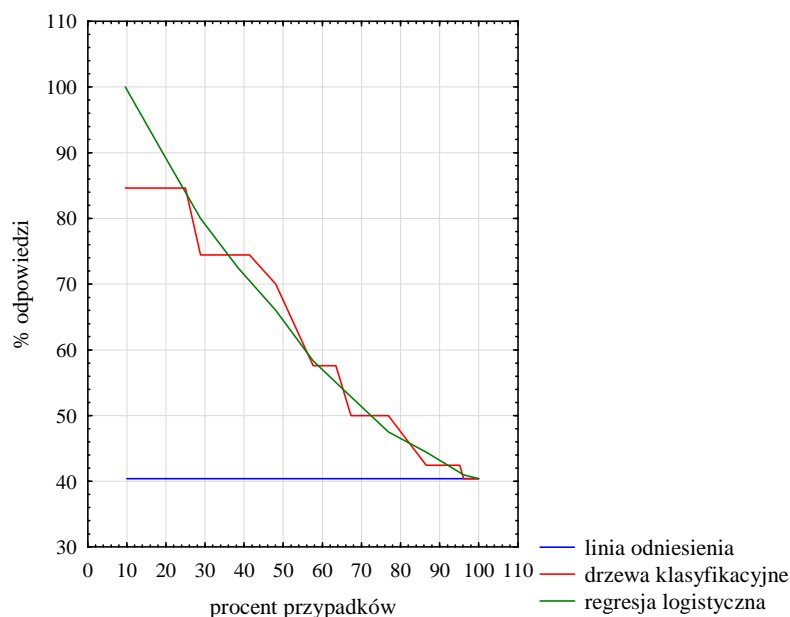
7.4.4 Osoby chorujące na nadciśnienie tętnicze i cukrzycę

Zmienne niezależne użyte do budowy modeli klasyfikujących osoby chorujące na nadciśnienie oraz cukrzycę i osoby z grupy kontrolnej to BMI oraz wskaźnik zmienności długoterminowej Zugaiba.

Tabela 48 Procent błędnych klasyfikacji

	drzewa klasyfikacyjne	regresja logistyczna
błąd	20,19%	23,08%

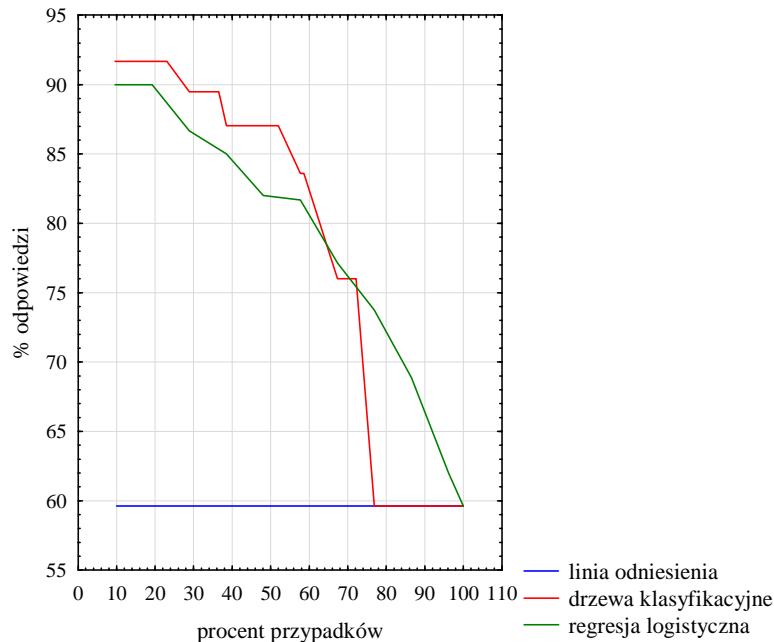
Regresja logistyczna cechuje się nieznacznie wyższym błędem klasyfikacji w porównaniu do modelu drzew klasyfikacyjnych. Procent błędnych klasyfikacji w modelu regresji jest o prawie 3 punkty procentowe wyższy w porównaniu do modelu drzew klasyfikacyjnych i wynosi 23,08 procent (**tabela 48**).



Wykres 74 Procent prawidłowych odpowiedzi dla osób chorujących na nadciśnienie tętnicze oraz cukrzycę

Dla osób chorujących na nadciśnienie tętnicze i cukrzycę, z wykresu 74 można odczytać, że biorąc 10 procent przypadków najpewniej zaklasyfikowanych do grupy badanej przez regresję logistyczną otrzymuje się próbę, w której 100 procent przypadków należy do grupy badanej. Analizując model drzew klasyfikacyjnych, na podstawie wykresu 74 można

zauważyć, że biorąc 20 procent przypadków najpewniej zaklasyfikowanych do grupy osób z nadciśnieniem tętniczym i cukrzycą, otrzymuje się próbę, w której 84 procent przypadków należy do tej grupy.



Wykres 75 Procent prawidłowych odpowiedzi dla osób z grupy kontrolnej

Z wykresu 75 można odczytać, że, w przypadku osób z grupy kontrolnej, biorąc 20% przypadków najpewniej przyporządkowanych przez drzewa klasyfikacyjne do grupy osób bez zawału mięśnia sercowego, nadciśnienia i cukrzycy otrzymujemy próbę, w której 92 procent przypadków należy do grupy kontrolnej. Natomiast w przypadku modelu regresji logistycznej biorąc 20% przypadków najpewniej zaklasyfikowanych do grupy kontrolnej otrzymujemy próbę, w której 90 procent przypadków należy do grupy osób bez zawału mięśnia sercowego, nadciśnienia i cukrzycy. Linia odniesienia odpowiada klasyfikacji losowej.

8. Dyskusja

Choroby sercowo - naczyniowe są jednymi z najczęstszych chorób dotykających ludzkość. Zalicza się do nich między innymi takie jednostki chorobowe jak: nadciśnienie tętnicze, choroba wieńcowa, zawał serca czy udar mózgu. W wielu przypadkach są także główną przyczyną przedwczesnej śmierci. Dlatego też analiza sygnału elektrokardiograficznego, w którym zmiany występują jeszcze przed pojawieniem się klinicznych objawów choroby, z punktu widzenia klinicysty, może być niezwykle przydatna. Stosowane są różne metody pomiaru zmienności rytmu serca: pomiary z zastosowaniem metod nieliniowych, analiza w dziedzinie czasu, czy analiza spektralna. Każda metoda bada zmienność rytmu serca pod innym kątem, dlatego też, mimo iż często ich wartości prognostyczne są zbliżone, niesie za sobą inną informację związaną z analizowanym sygnałem. Ponieważ zmienność rytmu serca może mieć wartość prognostyczną i rokowniczą w przypadku wielu schorzeń, dlatego badania nad rozwojem tej dziedziny medycyny trwają nadal. Pojawiają się nowe parametry określające zmienność rytmu serca, np. zdolność deceleracji rytmu serca (Deceleration Capacity, DC), opisujący ogólną zdolność do zwalniania rytmu serca [136]. Trwają również prace nad maksymalnym ograniczeniem obecności artefaktów i szumu podczas akwizycji sygnału, znacznie utrudniających jego analizę, np. poprzez zastosowanie filtrowania adaptacyjnego [137] oraz nad nowymi narzędziami ułatwiającymi wykonywanie samych analiz, a także prezentację graficzną wyników [138]. Analiza HRV znajduje zastosowanie w wielu zagadnieniach medycznych. Huikiri [139] przedstawił zastosowanie analizy zmienności rytmu serca u pacjentów po ostrym zawale mięśnia sercowego. Chiang i wsp. [140] badał związek pomiędzy wskaźnikami zmienności rytmu serca mierzonymi w dziedzinie częstotliwości a nieplanowanym przyjęciem do szpitala osób starszych. Chrysohoou i wsp. [141] badając powiązanie pomiędzy wskaźnikiem elastyczności aorty a zmiennością rytmu serca u starszych Ikaryjczyków wykazali, że obniżone wartości HRV korelują z utratą elastyczności przez aortę.

W pracy dokonano analizy danych rozważając cztery różne grupy badane. Różniły się one nie tylko analizowanymi jednostkami chorobowymi, ale również każda z nich miała różną liczebność. Pierwszą grupę stanowili pacjenci po przebytych zawale mięśnia sercowego, drugą – chorzy po przebytych zawale mięśnia sercowego i cierpiący dodatkowo na nadciśnienie tętnicze, trzecią - pacjenci po przebytych zawale mięśnia sercowego, chorujący na nadciśnienie tętnicze i cukrzycę. Ostatnia analizowana grupa badana składała się z osób

chorujących na nadciśnienie tętnicze i cukrzycę. Grupa kontrolna nie ulegała zmianie w wykonywanych analizach. Powtórzenie analiz czterokrotnie dla różnych grup badanych jest niewątpliwą zaletą pracy, gdyż pozwala na porównanie i wybór najistotniejszych parametrów zmienności rytmu serca w trzech wzajemnie przeplatających się jednostkach chorobowych. Dzięki temu wnioski uzyskane w pracy mogą być szerzej stosowane w praktyce.

Sygnal elektrokardiograficzny wykazuje zmienność zarówno w dziedzinie czasu jak i częstotliwości. Powszechnie stosowane parametry badające zmienność rytmu serca wyznaczone są albo bezpośrednio z odstępów RR albo opierają się na wartości HR, jednakże obie te wielkości wykazują zależność hiperboliczną.

Badanie własne oparto o analizę zarówno klasycznych parametrów zmienności rytmu serca mierzonych w dziedzinie czasu: SDNN i RMSSD, jak i wskaźników zmienności rytmu serca (zarówno krótko- jak i długoterminowej) wyznaczonych w oparciu o mniej powszechne modele matematyczne, przy czym wskaźniki zmienności długoterminowej ukazują zmiany częstości rytmu serca o niskiej częstotliwości, a krótkoterminowej – wysokiej częstotliwości. W celu wzbogacenia analiz o częstotliwościowe spojrzenie na zmienność rytmu serca, analizowano również mierzony w dziedzinie częstotliwości wskaźnik LF/HF, będący odzwierciedleniem równowagi współczulno-przywspółczulnej. Zbadano także wskaźnik Allana, będący reprezentantem parametrów zmierzonych w połączonych dziedzinach czasu i częstotliwości. Parametr ten nie zależy od funkcji HR, co jest jego niewątpliwą zaletą [31, 41, 142].

Dodatkowo do analizowanych czynników włączono wskaźnik wagowo-wzrostowy BMI oraz płeć i wiek pacjenta.

Interpretacja wyników dotyczących, płci, wieku i BMI

Obniżona zmienność rytmu serca mierzona przy pomocy klasycznych parametrów HRV ma związek m. in. z czynnikami osobniczymi oraz prowadzonym stylem życia. Wykazano na podstawie analizy spektralnej, że zmienność rytmu zatokowego jest ujemnie skorelowana z wiekiem pacjenta [65, 143, 144, 145]. W badaniach potwierdzono wpływ wieku, jako czynnika dyskryminacyjnego (w każdym analizowanym przypadku osoby z grupy badanej były istotnie starsze od osób z grupy kontrolnej). Voss i wsp. [146] analizując wpływ wieku

oraz płci na miary krótkoterminowej zmienności rytmu serca u 782 zdrowych kobiet i 1124 zdrowych mężczyzn, wysunęli przypuszczenie o znaczącej modyfikacji wartości analizowanych parametrów pod wpływem tych czynników, zwłaszcza w grupie młodszych pacjentów. Wiadomym jest również fakt, iż zdrowe, młodsze kobiety charakteryzują się wyższym tętnem oraz niższymi wartościami HRV w porównaniu do odpowiedniej wiekowo grupy zdrowych mężczyzn. Różnic tych nie zauważa się już w grupie osób starszych [147].

W pracy nie stwierdzono istnienia statystycznie istotnych różnic w poziomach analizowanych wskaźników zmienności rytmu serca pomiędzy kobietami i mężczyznami w żadnej grupie badanej. Znajduje to rzeczywiście potwierdzenie w powyżej przytoczonych pracach, gdyż analizowane w badaniu grupy stanowili pacjenci starsi (dla osób po przebytym zawale mięśnia sercowego średnia wieku wynosiła $59,2 \pm 9,7$, dla pacjentów po przebytym zawale i chorujących dodatkowo na nadciśnienie $60,5 \pm 10,4$, dla osób po przebytym zawale i cierpiących dodatkowo na nadciśnienie i cukrzycę $61,3 \pm 7,8$, a dla chorych na nadciśnienie i cukrzycę $60,6 \pm 9,4$). Wprawdzie Huikuri i wsp. [148] analizując grupę 50-ciolatków wykazali obniżenie niskoczęstotliwościowych składowych zmienności HRV oraz podwyższenie składowych wysokoczęstotliwościowych w grupie kobiet w porównaniu do grupy mężczyzn, jednak należy zwrócić uwagę, że kryterium włączenia do badania wykluczało pacjentów chorujących m. in. na nadciśnienie tętnicze, czy cukrzycę.

We wszystkich przeprowadzonych analizach, pacjenci z grup badanych byli istotnie starsi od pacjentów z grupy kontrolnej, co w świetle istniejących badań, może mieć istotny wpływ na obniżenie wartości HRV.

W przypadku współczynnika BMI badania dowodzą, iż u osób zdrowych wartości współczynnika BMI nie wpływają na spadek HRV [149]. Wykazano natomiast istnienie zależności pomiędzy otyłością a działaniem autonomicznego układu nerwowego [150]. Już 10-cio procentowy wzrost masy ciała powoduje wzrost HR. Obniżenie masy ciała powoduje natomiast zwolnienie częstości pracy serca [151].

W badaniach własnych pacjenci z analizowanych grup badanych mieli istotnie wyższe wartości wskaźnika BMI w porównaniu do osób z grupy kontrolnej, dla których mediana wartości tego współczynnika wynosiła 26,0 (min=20,0; maks=34,4). W każdej analizowanej grupie badanej mediana wartości BMI mieściła się w przedziale zdefiniowanym przez Światową Organizację Zdrowia jako nadwaga. Podobnie jak w przypadku wieku pacjenta, wartości współczynnika wagowo-wzrostowego mogą mieć wpływ na obniżenie wartości HRV.

Jedynie w przypadku osób po przebytych zawale mięśnia sercowego i chorujących dodatkowo na nadciśnienie tętnicze, nie wykazano istnienia istotnych różnic w wartościach wskaźnika BMI w stosunku do grupy kontrolnej.

Pojawia się zatem pytanie, czy obniżenie parametrów zmienności rytmu serca było skutkiem choroby czy raczej miało ono związek z czynnikami osobniczymi takimi jak wiek pacjenta czy jego BMI. Odpowiedzi na to pytanie starano się udzielić, rozważając jednoczesny wpływ wszystkich analizowanych zmiennych na zmienną zależną określającą wystąpienie danego schorzenia, przy zastosowaniu regresji logistycznej.

Interpretacja wyników uzyskanych dla klasycznych parametrów zmienności rytmu serca

W pracy wykazano obniżenie wartości parametru SDNN w porównaniu do grupy kontrolnej, zarówno w przypadku osób po przebytych zawale mięśnia sercowego, pacjentów po przebytych zawale mięśnia sercowego, chorujących dodatkowo na nadciśnienie tętnicze i cukrzycę, jak i osób chorujących na nadciśnienie i cukrzycę.

Należy pamiętać jednak o tym, że parametr SDNN wymaga długotrwałych zapisów analizy holterowskiej oraz niezwyklej staranności, by wykluczyć uderzenia ektopowe, brakujące oraz artefakty. Dlatego też należy podchodzić z rezerwą do uzyskanych wyników, które mogą być obciążone błędem wynikającym z krótkiej, 5-cio minutowej długości analizowanego sygnału.

W badaniach własnych u pacjentów po przebytych zawale mięśnia sercowego i chorujących dodatkowo na nadciśnienie i cukrzycę oraz osób cierpiących na nadciśnienie i cukrzycę zauważono istotnie niższe wartości współczynnika RMSSD w porównaniu do osób z grupy kontrolnej.

Podobnie, w przypadku wskaźnika RMSSD, który jest wrażliwy na duże różnice odstępów RR, długość analizowanego sygnału ma istotne znaczenie i interpretacja otrzymanych wyników powinna być przeprowadzona z dużą ostrożnością.

Wydaje się, że znaczący wpływ na powyższe wyniki ma fakt wystąpienia cukrzycy. Dotychczas prowadzone badania dowodzą obniżenia parametrów zmienności rytmu serca u pacjentów z cukrzycą w porównaniu do osób z właściwą glikemią [152, 153, 154, 155]. Baumert i Sacre [156], analizując grupę 16 pacjentów mających zdiagnozowaną cukrzycę

typu 2, zauważyli wzrost spoczynkowej częstości akcji serca i obniżenie krótkoterminowych wskaźników zmienności rytmu serca. Roy i Ghatak [157], wykorzystując wykres Poincaré do porównania zmienności rytmu serca u pacjentów z cukrzycą typu 2, wykazali obniżenie modulacji przywspółczulnej. Podobnie, znacząco niższe wartości parametrów mierzonych w dziedzinie czasu i częstotliwości mieli pacjenci chorujący na cukrzycę typu 1 w porównaniu do osób z grupy kontrolnej w badaniu przeprowadzonym przez Turkera [158].

W przypadku nadciśnienia tętniczego sytuacja nie jest tak jednoznaczna. Uzyskane w badaniach własnych wyniki świadczące o statystycznie istotnie niższych wartościach wskaźników HRV są zgodne z wynikami uzyskanymi przez Radelli [83] i Huikuri [84].

Badania Framingham Heart Study [79] i Atherosclerosis Risk In Communities (ARIC) [80] również wykazały związek obniżonej całkowitej HRV z zapadalnością na choroby układu sercowo-naczyniowego. Jednakże badania przeprowadzone przez Mancina i wsp. nie potwierdziły obniżenia wartości klasycznych parametrów zmienności rytmu serca u pacjentów z nadciśnieniem łagodnym i ciężkim. Należy pamiętać jednak o tym, iż wymienione badania różniły się znacznie pod względem liczebności badanych grup. Występowały w nich również niejednorodności, co mogło mieć znaczący wpływ na otrzymanie niespójnych wniosków.

Obniżenie wartości wskaźników zmienności rytmu serca u pacjentów po przebytych zawale mięśnia sercowego wykazane w powyższej pracy, znajduje potwierdzenie w badaniach Osterheusa i wsp. [159]. Także Oliveira i wsp. [160] badając wpływ ćwiczeń fizycznych na zmienność rytmu serca u pacjentów po zawale mięśnia sercowego, wskazują na obniżenie HRV, co ma bezpośrednie przełożenie na rosnące ryzyko śmierci.

Interpretacja wyników uzyskanych dla parametrów zmienności rytmu serca wyznaczonych w oparciu o rzadziej stosowane modele matematyczne, wskaźnik LF/HF oraz wskaźnik Allana

W ocenie diagnostycznej czy rokowniczej większą rolę odgrywa zmienność długoterminowa (z klinicznego punktu widzenia). Słusznie zatem może pojawić się wątpliwość, czy możliwa jest analiza zmienności długoterminowej w oparciu o krótkie fragmenty zapisu sygnału elektrokardiograficznego. Definicje zastosowanych w pracy

wskaźników zmienności rytmu serca, do oceny zmienności długoterminowej, wykorzystują stosunkowo niewielką liczbę odstępów RR i są zbudowane w taki sposób, by uchwycić rzeczywistą zmienność, a nie chwilowe oscylacje. Zastosowane algorytmy zostały przez autorów dobrane w taki sposób, by jak najlepiej przybliżyć interpretację sygnałów biologicznych. Zastosowanie przez Zugaiba mediany, która jest mniej czuła na wartości skrajne w porównaniu do średniej arytmetycznej, jest tego najlepszym przykładem.

Z racji tego, iż bezpośrednie porównywanie między sobą właściwości poszczególnych wskaźników zmienności rytmu serca jest niemożliwe, co wynika przede wszystkim z różnicy w wymaganej szerokości okna czasowego, w którym rejestrowany jest sygnał, w pracy zastosowano standaryzację uwzględniając parametry obliczone w znormalizowanym jednonminutowym oknie czasowym.

W badaniu nie analizowano parametrów HRV uzyskanych metodą geometryczną, gdyż wyrażają one zmienność całkowitą mierzona w ciągu 24 h [161] i nie mogą być stosowane do oceny zmienności krótkoterminowej.

Przeprowadzone analizy jednowymiarowe, porównujące poziomy analizowanych parametrów pomiędzy grupą badaną i kontrolną potwierdziły doniesienia wskazujące na obniżenie wartości niektórych parametrów zmienności rytmu serca u pacjentów z grupy badanej.

U osób po przebytych zawale mięśnia sercowego wykazano obniżenie wartości wskaźników zmienności rytmu serca mierzonych w dziedzinie czasu, zarówno krótko- jak i długoterminowych. Osoby chore miały niższe wartości wskaźników: de Haana, Yeha, van Geijna, Hueya, Daltona i Zugaiba, zarówno krótko- jak i długoterminowych.

W grupie osób po przebytych zawale mięśnia sercowego i chorujących dodatkowo na nadciśnienie tętnicze stwierdzono istotne obniżenie wartości współczynników zmienności krótkoterminowej de Haana, van Geijna i Daltona i długoterminowej de Haana i Zugaiba.

Analizując grupę chorych po przebytych zawale mięśnia sercowego, cierpiących dodatkowo na nadciśnienie i cukrzycę wykazano istnienie istotnych statystycznie różnic w wartościach większości współczynników zmienności krótko- i długoterminowej, zmierzonych w dziedzinie czasu oraz w wartościach parametru zmierzonego w połączonych dziedzinach czasu i częstotliwości – wskaźnika Allana. Nie zaobserwowano różnic w wartościach zmierzonego w dziedzinie czasu wskaźnika zmienności długoterminowej Organa.

Osoby chorujące na nadciśnienie tętnicze i cukrzycę miały istotnie niższe wartości większości współczynników zmienności krótko- i długoterminowej, zmierzonych w dziedzinie czasu oraz parametru zmierzonego w połączonych dziedzinach czasu i częstotliwości – wskaźnika Allana w porównaniu do osób z grupy kontrolnej.

Pawlak-Buś [162] analizując pięciominutowe zapisy elektrokardiografu pacjentów znajdujących się w pozycji pionowej, dowodzi obniżenia wartości wskaźników zmienności długoterminowej Yeha, Hueya, Zugaiba i Organa, krótkoterminowej Hueya oraz parametru moc całkowita (Total Power, TP) u pacjentów z klinicznymi objawami choroby wieńcowej w porównaniu do osób zdrowych.

W grupach pacjentów po przebytych zawałach mięśnia sercowego, chorujących na nadciśnienie tętnicze i cukrzycę oraz osób chorujących na nadciśnienie tętnicze i cukrzycę zauważono najwięcej istotnych różnic w poziomach analizowanych parametrów w porównaniu do grupy kontrolnej (w pierwszym przypadku różnice wykazano w poziomach 15 parametrów, w przypadku drugiej grupy – 16 zmiennych). Może to świadczyć o tym, że współistnienie nadciśnienia tętniczego wraz z cukrzycą wzmacnia zakłócenia równowagi pomiędzy częścią współczulną i przywspółczulną AUN. Również tylko w tych grupach zaobserwowano różnicę w poziomie wskaźnika Allana, który jest reprezentantem parametrów mierzonych w połączonych dziedzinach czasu i częstotliwości, odnotowując istotny jego wzrost w porównaniu do grupy kontrolnej.

W żadnej z analizowanych grup nie stwierdzono istnienia statystycznie istotnych różnic w poziomie częstotliwościowego parametru LF/HF. Zatem pojawia się pytanie o celowość uwzględniania tego wskaźnika w dalszych analizach. Okazuje się jednak, że niektóre badania, zwłaszcza dotyczące cukrzycy, potwierdzają występowanie różnic w poziomie LF/HF pomiędzy chorymi i zdrowymi. Turker [158], analizując 35 chorych ze zdiagnozowaną cukrzycą typu 1 wykazał obniżenie wartości wskaźnika LF/HF w stosunku do wartości tego parametru zmierzonego w 35-cio osobowej grupie kontrolnej. Nie znajduje to jednak potwierdzenia w badaniach Pal i wsp. [163], którzy wykazali znaczące podwyższenie parametru LF/HF w grupie osób zdrowych, którzy posiadali krewnych pierwszego stopnia ze zdiagnozowaną cukrzycą typu 2 w porównaniu do grupy kontrolnej. Ze względu na to, iż badania były prowadzone na różnych grupach (chorzy w przypadku Turkera i zdrowi z dodatnim wywiadem rodzinnym u Pal), dlatego też nie można ich porównywać. Najlepszym rozwiązaniem wydaje się więc przeprowadzenie szerszych badań dotyczących zmienności rytmu serca w ujęciu częstotliwościowym.

Zastosowane metody

Zainteresowanie badaczy poszukiwaniem wzajemnych powiązań pomiędzy analizowanymi zmiennymi i jednoczesnym ich wpływem na zmienną zależną skłania do poszukiwań metod, pozwalających na wykonywanie takich wielowymiarowych analiz.

W pracy zaprezentowano porównanie wyników klasyfikacji otrzymanych w wyniku zastosowania dwóch zupełnie różnych metod: statystycznej – regresji logistycznej oraz DATA MINING – drzew klasyfikacyjnych.

Modele regresji logistycznej znajdują zastosowanie w objaśnianiu jednoczesnego wpływu zmiennych niezależnych na zmienną zależną, która najczęściej jest zmienną dychotomiczną. Metoda ta jest coraz chętniej stosowana w praktyce, co wynika z tego, iż jej zastosowanie nie wymaga spełnienia tak wielu założeń dotyczących zmiennych niezależnych jak w przypadku liniowej regresji wielokrotnej.

Zmienne niezależne budujące model mogą być zarówno dyskretne jak i ciągłe. Nie muszą także spełniać warunku homoscedastyczności. Należy pamiętać jednak o tym, by ich liczba w modelu nie była zbyt duża (sugeruje się, by przestrzegać warunku $n > 10(k+1)$, gdzie k – liczba zmiennych niezależnych, n – licznosc próby). Regresję logistyczną można stosować także dla prób o niewielkiej liczebności. Dzięki tej metodzie można wykrywać zjawiska globalne, a wyniki przedstawione za pomocą ilorazu szans są łatwe w interpretacji.

W pracy, w zależności od analizowanej grupy badanej, zbudowano różną liczbę modeli logistycznych, opartych o różne kombinacje współczynnika BMI, wieku oraz wskaźników zmienności rytmu serca. Zmienna płeć nie uzyskała istotności statystycznej w żadnym z budowanych modeli. Wiek był czynnikiem istotnym w klasyfikacji pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie.

Dla pacjentów po przebytych zawale mięśnia sercowego oraz osób chorujących na nadciśnienie i cukrzycę, zbudowano cztery różne modele przyporządkowujące osoby z grupy badanej i kontrolnej do właściwej klasy. Dla pozostałych dwóch grup zbudowano tylko jeden model logistyczny.

W przypadku osób po przebytych zawale mięśnia sercowego, do dalszej analizy wybrano model, który miał najwyższy iloraz szans, największy procent poprawnie zaklasyfikowanych przypadków, najwyższą czułość i swoistość w porównaniu do pozostałych modeli. Zbudowano go w oparciu o zmienną BMI i współczynnik zmienności długoterminowej de Haana.

Dla pacjentów chorujących na nadciśnienie i cukrzycę do dalszej analizy wybrano model zbudowany w oparciu o współczynnik BMI i wskaźnik zmienności długoterminowej Zugaiba. Ma on gorsze parametry w porównaniu do modelu zbudowanego o BMI, wskaźniki de Haana, Yeha, van Geijna i Zugaiba (mniejszy iloraz szans, niższy procent poprawnie zaklasyfikowanych przypadków). Jednak analizując budowę modelu opartego o 7 powyższych zmiennych można zauważyć, że prawie wszystkie zmienne włączone do modelu są ze sobą silnie skorelowane. W niektórych przypadkach współczynniki korelacji rangowej przyjmują wartości większe niż 0,9, co świadczy o istnieniu wręcz bardzo silnej zależności monotonicznej (**Tabela A11 Załącznik 2**). Cztery z włączonych zmiennych są nieistotne statystycznie i pełnią rolę wzmacniającą dla pozostałych zmiennych (usunięcie tych nieistotnych statystycznie parametrów z modelu skutkowało utratą istotności statystycznej całego modelu) **Załącznik 2**.

Wybierając zmienne do modelu należy uwzględnić główne założenie regresji logistycznej, mówiące o tym, że parametry wchodzące w skład modelu powinny być liniowo niezależne (nie powinno mieć miejsca zjawisko współliniowości) [135]. Włączenie zmiennych skorelowanych zmniejsza szansę na wykrycie badanej zależności, gdyż powoduje zwiększenie przedziałów ufności, budowanych dla odpowiednich zmiennych.

W praktyce jednak większość danych pochodzących od jednego obiektu jest ze sobą skorelowana. Najczęściej przypadek niewspółliniowości ma miejsce dla danych eksperymentalnych. Dlatego też zakłada się, że zmienne włączane do modelu logistycznego nie powinny być redundantne, tzn. że każda zmienna powinna wносить do modelu nowe informacje. Często dochodzi do sytuacji, w której pomiary są ze sobą silnie skorelowane, ale nie jest to zależność liniowa. Może to prowadzić np. do sytuacji, w której niewielka zmiana w zbiorze danych spowoduje duże zmiany w estymatorach. Silna zależność pomiędzy parametrami modelu może powodować odwrócenie kierunku „wpływu” zmiennej niezależnej na zmienną zależną. Mówi się wówczas o występowaniu efektu interakcji między zmiennymi, tzn. że wpływ zmiennej niezależnej na zmienną zależną zmienia się pod wpływem innej zmiennej, zwanej moderatorem. W przypadku regresji logistycznej moderatorem jest zmienna niezależna, której wartości powodują zmianę siły lub/i kierunku analizowanej zależności [164, 165].

Taka sytuacja ma miejsce w przypadku modelu zbudowanego w celu klasyfikacji osób po przebytym zawale mięśnia sercowego i chorujących na nadciśnienie. Współczynnik zmienności krótkoterminowej de Haana i długoterminowej Organa mają dodatnie wartości oceny parametrów oraz ilorazy szans większe niż 1, co sugerowałoby, że wzrost wartości

tych zmiennych zwiększa ryzyko wystąpienia analizowanej jednostki chorobowej. Jest to sprzeczne z wynikami uzyskanymi w analizie jednowymiarowej i z doniesieniami literaturowymi, mówiącymi o obniżeniu wartości HRV na skutek rozważanych chorób. Za uzyskanie takich wyników odpowiedzialne jest włączenie do modelu zmiennych silnie skorelowanych (**Tabela A9 Załącznik 2**). Współczynniki korelacji rangowej pomiędzy wskaźnikami zmienności krótkoterminowej de Haana, Yeha oraz van Geijna, parametrami zmienności długoterminowej Organa i Zugaiba oraz pomiędzy wartościami wskaźnika zmienności krótkoterminowej van Geijna i długoterminowej Zugaiba przyjmują wartości większe od 0,7, co wskazuje na istnienie silnej korelacji pomiędzy tymi zmiennymi.

Celowo nie pominięto tych parametrów podczas budowania modelu, z racji tego iż usunięcie z modelu predyktora podejrzanego o powodowanie współliniowości byłoby jednoznaczne z przyjęciem hipotezy, że dany parametr nie wpływa na wystąpienie choroby. Natomiast błąd selekcji zmiennych może wywołać gorsze skutki niż występowanie współliniowości [166]. W badaniach własnych pominięcie którejkolwiek z silnie skorelowanych zmiennych powodowało utratę istotności pozostałych zmiennych wchodzących w skład modelu i tym samym całego modelu, co potwierdza, że ma on istotny wpływ na wystąpienie zawału mięśnia sercowego przy współistnieniu nadciśnienia.

Sposobem rozwiązania powyższego problemu mogłoby być rozważenie efektów interakcji pomiędzy poszczególnymi zmiennymi i włączenie ich do modelu, co może stanowić kierunek dalszych badań.

We wszystkich analizach istotnym parametrem okazał się być współczynnik wagowo - wzrostowy (iloraz szans wynosił od 1,16 do 1,42 w zależności od analizowanego schorzenia, przedziały ufności nie zawierały 1). Ponieważ można go w bardzo prosty sposób obliczyć, więc może znaleźć powszechne zastosowanie w praktyce lekarskiej w celu oceny ryzyka wystąpienia którejkolwiek z rozważanych chorób.

Wyniki uzyskane dla pozostałych modeli są adekwatne do wyników uzyskanych w analizach jednowymiarowych i pokrywają się z doniesieniami literaturowymi. Wzrost wskaźnika BMI oraz obniżenie wartości istotnych powodują zwiększenie ryzyka wystąpienia rozważanych jednostek chorobowych. Dodatkowo, w przypadku osób po przebytym zawałe mięśnia sercowego i chorujących na nadciśnienie, obok BMI i wskaźników zmienności rytmu serca, czynnikiem istotnym jest wiek pacjenta, którego wzrost powodował zwiększenie ryzyka wystąpienia rozważanej jednostki chorobowej.

Dla wszystkich modeli, w celu sprawdzenia jak potrafią klasyfikować nowe, nieznane przypadki, przeprowadzono walidację w oparciu o zbiór uczący i testowy. Wyniki uzyskane

na zbiorze testowym są prawie tak dobre jak na zbiorze uczącym, co świadczy o tym, że modele dobrze klasyfikują nowe obiekty i mogą być z powodzeniem stosowane w praktyce.

Z racji tego, iż celem pracy było porównanie dwóch różnych metod służących klasyfikacji danych, przyjęto założenie, że obie techniki powinny być oparte o te same zmienne. Dzięki temu założeniu można było dodatkowo zastosować metodę szybkiego wdrażania modeli predykcyjnych, która służy do oceny skuteczności omawianych modeli. W celu porównania wyników klasyfikacji otrzymanych przy zastosowaniu regresji logistycznej i drzew klasyfikacyjnych uznano, że należy zestawić wyniki uzyskane na całym zbiorze danych. Rozbieżności w liczebnościach wynikają z zastosowania różnych metod walidacji: w przypadku regresji logistycznej zastosowano podział na zbiór uczący i testowy, w przypadku drzew klasyfikacyjnych globalny sprawdzian krzyżowy. W tabelach 43 i 44 przedstawione zostały wyniki klasyfikacji otrzymane przez regresję logistyczną na całym zbiorze danych i przez drzewa klasyfikacyjne w globalnym sprawdzianie krzyżowym, który również uwzględnia tę samą liczbę przypadków.

Niestety, modele zbudowane w celu przyporządkowania osób chorujących na nadciśnienie i cukrzycę oraz osób z grupy kontrolnej do właściwej klasy miały podobne parametry klasyfikacyjne, co uniemożliwiło wybór najlepszego. Zastosowanie szybkiego wdrażania modeli predykcyjnych również nie rozwiązało powyższego problemu.

W przypadku osób po przebytym zawale mięśnia sercowego, dodatkowo chorujących na nadciśnienie oraz osób po przebytym zawale i chorujących na nadciśnienie i cukrzycę modelem, który skuteczniej przyporządkowywał obiekty do klas okazała się być regresja logistyczna.

Drzewa klasyfikacyjne były skuteczniejsze w klasyfikacji pacjentów po przebytym zawale mięśnia sercowego.

Sowińska [167], porównując różne modele klasyfikujące dzieci z nadciśnieniem i z prawidłowym ciśnieniem do odpowiedniej klasy, wykazała, że model daje najlepszą klasyfikację, gdy jest zbudowany w oparciu o „własne zmienne”, czyli predyktory wybrane do jego budowy przez zaimplementowane algorytmy. Na tej podstawie można wysnuć przypuszczenie, że klasyfikacja uzyskana przez drzewa klasyfikacyjne mogłaby być lepsza, gdyby dopuścić możliwość swobodnego wyboru predyktorów przez drzewo. Ponieważ różnice w otrzymanych wartościach klasyfikacji są niewielkie, można przyjąć hipotezę, że drzewa mogłyby się okazać metodą bardziej efektywną w klasyfikacji analizowanych jednostek chorobowych. Dlatego też ciekawym wydaje się pomysł zbadania jakie wartości

klasyfikacyjne uzyskałyby drzewa klasyfikacyjne, gdyby analizy zostały przeprowadzone na zmiennych wybranych przez drzewo i porównanie tych wartości z parametrami uzyskanymi przy zastosowaniu regresji logistycznej

Wyniki przedstawione w pracy dowodzą, że drzewa klasyfikacyjne mogą być dobrą alternatywą dla badaczy zajmujących się problemami klasyfikacji, zwłaszcza w sytuacjach, gdy zastosowanie tradycyjnych metod statystycznych (np. regresji logistycznej) jest niemożliwe ze względu na brak spełnienia założeń. Jednakże często zdarza się, że modele zaproponowane przez drzewa klasyfikacyjne są trudno interpretowalne. Wynika to najczęściej z sugerowanych kryteriów podziałów przypisanych do konkretnego węzła, które nie znajdują potwierdzenia w przesłankach klinicznych. Statistica 10 daje możliwość zmiany tych mało intuicyjnych kryteriów na takie, które dla badacza będą miały sens, poprzez moduł interakcyjnych drzew klasyfikacyjnych.

Wykresy A1–A4 (Załącznik 4) przedstawiają strukturę drzew klasyfikacyjnych zbudowanych dla poszczególnych jednostek chorobowych w oparciu o powyższą aplikację. Kryteria podziału dla zmiennej BMI zostały zmienione w taki sposób, by były to liczby całkowite. Dzięki temu będzie je można łatwiej odnieść do klasyfikacji BMI zaproponowanej przez Światową Organizację Zdrowia. Wiek podano z dokładnością do pół roku. Po konsultacji z kardiologami ustalono, iż z powodu braku jakichkolwiek kryteriów podziału dla wskaźników zmienności rytmu serca, kryteria podziałów dla tych zmiennych zostaną zaokrąglone z dokładnością do dwóch lub trzech miejsc po przecinku, w zależności od rzędu wielkości analizowanych parametrów.

W **tabelach A18, A20, A22 i A24 (Załącznik 4)** przedstawiono klasyfikację przypadków z grupy badanej i kontrolnej, odpowiednio do analizowanej grupy.

Tabele A19, A21, A23 i A25 (Załącznik 4) zawierają porównanie wyników klasyfikacji: procentu poprawnie zaklasyfikowanych przypadków, czułości i swoistości, dla drzew klasyfikacyjnych zbudowanych w oparciu o kryteria podziału zaproponowane przez drzewo oraz dla drzewa o zmienionych, wg powyżej opisanych zasad, kryteriach podziału.

Zmiana kryteriów podziału spowodowała nieznaczny spadek poprawności klasyfikacji pacjentów po przebytym zawale mięśnia sercowego i osób z grupy kontrolnej w porównaniu do klasyfikacji przeprowadzonej w oparciu o pierwotne kryteria podziału. Model ten gorzej

klasyfikuje osoby z grupy kontrolnej (ma niższą swoistość). Zmiana kryteriów podziału spowodowała natomiast, iż nieznacznie wzrosła czułość zaproponowanego modelu.

Podobnie jak w przypadku osób po przebytych zawale mięśnia sercowego, zmiana kryteriów podziału spowodowała nieznaczny spadek poprawności klasyfikacji pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze oraz osób z grupy kontrolnej w porównaniu do klasyfikacji przeprowadzonej w oparciu o pierwotne kryteria podziału. Model zbudowany w oparciu o zmodyfikowane kryteria gorzej klasyfikuje osoby z grupy badanej (ma niższą czułość). Cechuje go również nieco niższa swoistość.

Dokonana zmiana kryteriów podziału nie spowodowała żadnych zmian w klasyfikacji pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę oraz osób z grupy kontrolnej w porównaniu do klasyfikacji przeprowadzonej w oparciu o pierwotne kryteria podziału.

W przypadku klasyfikacji pacjentów chorujących na nadciśnienie tętnicze i cukrzycę i osób z grupy kontrolnej, zmiana kryteriów podziału spowodowała nieznaczny spadek poprawności klasyfikacji w porównaniu do klasyfikacji przeprowadzonej w oparciu o pierwotne kryteria podziału (o niecały punkt procentowy). Model ten gorzej klasyfikuje osoby z grupy kontrolnej (ma niższą swoistość). Zmiana kryteriów podziału spowodowała, iż wzrosła czułość zaproponowanego modelu.

Zmiana kryteriów podziału powodowała niewielkie pogorszenie wyników klasyfikacji lub, jak w przypadku grupy pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie i cukrzycę, nie zmieniła jej wartości, czy wreszcie, jak w przypadku osób po przebytych zawale mięśnia sercowego oraz osób chorujących na nadciśnienie tętnicze i cukrzycę, podniosła czułość metody. Dlatego też model zbudowany w oparciu o zmodyfikowane kryteria podziału może znaleźć zastosowanie w rozwiązywaniu problemów klasyfikacyjnych, jako łatwiejszy w interpretacji i bardziej intuicyjny w porównaniu do modelu zbudowanego w oparciu o pierwotne kryteria.

Pomimo iż w większości analizowanych problemów regresja logistyczna okazała się skuteczniejszą metodą rozwiązującą problem klasyfikacji osób z grupy badanej i kontrolnej, to niewielkie różnice w uzyskanych wynikach klasyfikacji dla drzew klasyfikacyjnych mogą być nadzieją dla badaczy, którzy ze względu na brak spełnienia założeń nie mogą zastosować tradycyjnych metod analizy danych.

Wyników uzyskanych w niniejszej pracy nie należy traktować jako ścisłych wniosków kardiologicznych. Intencją było opracowanie bazy metodologicznej do projektowania przyszłych systemów wspomaganie diagnostyki.

9. Wnioski

1. W przypadku osób po przebytych zawale mięśnia sercowego dodatkowo chorujących na nadciśnienie oraz osób po przebytych zawale mięśnia sercowego cierpiących dodatkowo na nadciśnienie i cukrzycę modelem, który najdokładniej opisywał analizowane zagadnienie był model regresji logistycznej. W porównaniu do modelu zaproponowanego przez drzewa klasyfikacyjne cechował się on wyższym procentem poprawnie zaklasyfikowanych przypadków. Miał również wyższą swoistość i był bardziej czuły.

W przypadku osób po przebytych zawale mięśnia sercowego model drzew klasyfikacyjnych okazał się skuteczniejszy. Cechował się nieco wyższym procentem poprawnie zaklasyfikowanych przypadków. W porównaniu do modelu regresji logistycznej miał wprawdzie nieco niższą czułość, lecz zdecydowanie wyższą swoistość. Potwierdziła to zastosowana do oceny jakości modeli regresji logistycznej i drzew klasyfikacyjnych metoda DATA MINING szybkiego wdrażania modeli predykcyjnych. Regresja logistyczna cechowała się wyższym błędem klasyfikacyjnym. Dla przypadków, którym modele dawały największe prawdopodobieństwo klasyfikacyjne, drzewa klasyfikacyjne dawały największą poprawność klasyfikacji zarówno dla przypadków z grupy badanej jak i kontrolnej.

Modele zbudowane w celu opisanie osób chorujących na nadciśnienie i cukrzycę oraz osób z grupy kontrolnej, mają podobny odsetek poprawnie zaklasyfikowanych przypadków. Dlatego nie można jednoznacznie wskazać, która metoda rzetelniej odzwierciedla analizowany problem. Wynika to z tego, iż model regresji logistycznej cechuje się niższą czułością oraz wyższą swoistością w porównaniu do modelu drzew klasyfikacyjnych. Dlatego też, decyzję wyboru metody należy pozostawić badaczowi.

2. W charakterystyce analizowanych jednostek chorobowych największe znaczenie ma współczynnik wagowo – wzrostowy BMI, gdyż pojawiał się jako istotny czynnik we wszystkich analizowanych modelach. Dodatkowo, wiek miał istotne znaczenie przy opisywaniu osób po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie oraz pacjentów z grupy kontrolnej.

Wśród wskaźników zmienności rytmu serca najistotniejsze okazały się: wskaźnik zmienności krótkoterminowej de Haana oraz wskaźniki zmienności długoterminowej de

Haana i Zugaiba. Dla osób chorujących na zawał oraz dla osób cierpiących na nadciśnienie i cukrzycę, obok BMI, najistotniejszy okazał się współczynnik zmienności długoterminowej. Dla osób z zawałem, nadciśnieniem i cukrzycą istotnym parametrem, obok BMI, był współczynnik zmienności krótkoterminowej. Dla chorych po przebytych zawałach mięśnia sercowego i cierpiących dodatkowo na nadciśnienie tętnicze, obok wieku i BMI, tak samo istotne okazały się być zarówno wskaźniki zmienności krótko jak i długoterminowej.

10. Streszczenie

W pracy porównano dwie metody: regresję logistyczną i drzewa klasyfikacyjne, w celu znalezienia modelu, który najlepiej charakteryzuje osoby po przebytych zawale mięśnia sercowego, chorujące na nadciśnienie tętnicze lub/i cukrzycę. Celem badania było znalezienie zmiennych, które mają najistotniejsze znaczenie w opisywaniu analizowanych jednostek chorobowych. W analizach uwzględniono wskaźniki zmienności rytmu serca zmierzone w dziedzinie czasu, częstotliwości (LF/HF) oraz w połączonych dziedzinach czasu i częstotliwości (wskaźnik Allana). Wśród parametrów zmierzonych w dziedzinie czasu znalazły się zarówno klasyczne parametry (SDNN, RMSSD) oraz wskaźniki mniej powszechne, określające zmienność zarówno krótko jak i długoterminową: de Haana, Yeha, van Geijna, Hueya, Daltona, Zugaiba oraz szerokość wstęgi oscylacyjnej. Dodatkowo zbadano wpływ takich zmiennych jak płeć, wiek czy BMI, jako zmiennych wnikających w modelach.

W badaniu wzięło udział 307 osób zakwalifikowanych do planowego zabiegu koronarografii w stabilnym okresie choroby niedokrwiennej serca. Analizy przeprowadzono w czterech wybranych grupach pacjentów: po przebytych zawale mięśnia sercowego, po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze, po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę oraz w grupie pacjentów cierpiących na nadciśnienie tętnicze oraz cukrzycę.

W charakterystyce pacjentów po przebytych zawale mięśnia sercowego i osób z grupy kontrolnej, techniką, która okazała się być najbardziej skuteczną były drzewa klasyfikacyjne. Dla pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie oraz osób po przebytych zawale mięśnia sercowego i cierpiących dodatkowo na nadciśnienie i cukrzycę modelem, który najlepiej opisywał analizowane zagadnienie okazał się model regresji logistycznej, cechujący się najwyższą czułością i swoistością.

Nie udało się wybrać optymalnego modelu dla pacjentów chorujących na nadciśnienie tętnicze i cukrzycę oraz osób z grupy kontrolnej.

W opisywaniu analizowanych jednostek chorobowych największe znaczenie miał współczynnik BMI, gdyż pojawiał się jako istotny czynnik we wszystkich analizowanych modelach. Spośród analizowanych wskaźników zmienności rytmu serca najistotniejsze okazały: wskaźnik zmienności krótkoterminowej de Haana oraz wskaźniki zmienności długoterminowej de Haana i Zugaiba.

Dla osób chorujących na zawał oraz dla osób cierpiących na nadciśnienie i cukrzycę, obok BMI, najistotniejszy okazał się współczynnik zmienności długoterminowej.

Dla osób z zawałem, nadciśnieniem i cukrzycą istotnym parametrem, obok BMI, był współczynnik zmienności krótkoterminowej.

Dla chorych po przebytych zawałach mięśnia sercowego i cierpiących dodatkowo na nadciśnienie tętnicze, obok BMI i wieku, tak samo istotne okazały się być zarówno wskaźniki zmienności krótko jak i długoterminowej.

11. Summary

The subject area of presented doctoral thesis is comparison of two mathematical models: logistic regression and classification trees in order to find the one that best describes a person after myocardial infarction with hypertension and / or diabetes. The aim is to find variables that can be used in the characteristics of the analyzed diseases. Heart rate variability indicators, measured in time, frequency (LF/HF) and joint time and frequency (the Allan factor) domains are considered. Moreover, gender, age, and BMI index are analyzed as confounding variables in the model.

Classical parameters (SDNN, RMSSD) and both long and short term HRV indicators: de Haan, Yeh, van Geijn, Huey, Dalton, Zugaib and oscillation index are analyzed.

The research group consists of 307 patients qualified for elective coronary angiography in stable coronary heart disease. Four different groups are explored: patients after myocardial infarction, patients after myocardial infarction with hypertension, patients after myocardial infarction with hypertension and diabetes and patients with hypertension and diabetes.

For patients after myocardial infarction, classification trees turned out to be the best technique.

Logistic regression model, with the highest sensitivity and specificity, is the best to describe patients after myocardial infarction with hypertension and patients after myocardial infarction with hypertension and diabetes and the control group's patients.

The study failed to find the optimal model for patients with hypertension and diabetes.

The most significant variable which was chosen of those analyzed appeared to be the BMI index. This variable is significant in all analyzed models. The most important HRV parameters turned to be de Haan's short term indicator and long term de Haan's and Zugaib's index.

The most significant variables appeared to be appropriately: for patients after myocardial infarction and for patients with hypertension and diabetes BMI index and long term variability parameter, for patients after myocardial infarction with hypertension and diabetes – BMI index and short term variability parameter. For patients after myocardial infarction with hypertension, beside BMI index and age, long term variables were as important as short term variability indicators.

12. Bibliografia

1. Cygankiewicz I., Zareba W. Heart rate variability, Handbook of Clinical Neurology, 2013; 117: 379-93
2. Hon EH, Lee ST. Electronic evaluations of the fetal heart rate patterns preceding fetal death, further observations. Am J Obstet Gynae 1965; 87:814–26
3. Roy B, Ghatak S. Nonlinear Methods to Assess Changes in Heart Rate Variability in Type 2 Diabetic Patients, Arquivos Brasileiros de Cardiologia, 2013 Oct;101(4):317-27
4. Hales S. Haemastatics In: Hales S. (ed), Statistical Essays, vol II, Innings, 1733; Manby & Woodward, London
5. Hon E.H., Lee S.T. Electronic evaluations of the fetal heart rate patterns preceding fetal Heath: further observations, Am. J. Obstet. Ginecol. 1965; 87:814
6. Murray A., Ewing D.J., Campbell I.W., Neilson J.M.M., Clarke B.F. RR interval variations in young male diabetics, Br Heart J., 1975; 37: 882- 885
7. Ewing D.J., Martin C.N., Young R.J., Clarke B.F. The value of cardiovascular autonomic function tests: 10 years experience in diabetes, Diabetic Care, 1985; 8:491 -498
8. Wolf M.M., Varigos G.A., Hunt D., Sloman J.G., Sinus arrhythmia in acute myocardial infarction, Med. J. Aust. 1978; 2:52
9. Kleiger R.E., Miller J.P., Bigger J.T., Moss A.J., Decreased heart rate variability and its association with increased mortality after acute myocardial infarction, Am. J. Cardiol., 1987; 59: 256 – 262
10. Bigger J.T., Kleiger R.E., Fleiss J.L., Components of heart rate variability measured during healing of acute myocardial infarction., Am. J. Cardiol., 1988;61:208-215
11. Causes of death in the EU, Statistics in focus 10/2006,

http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-NK-06-010

12. Pawlak-Buś K., Kołodziejczyk-Feliksik M., Kramer L., Nikisch E., Moczko J., Siminiak T., The Allan factor: a New model of mathematical interpretation of heart rate variability in stable coronary artery disease. Preliminary results., *Kardiol. Pol.*, 2005; 63: 125 – 132

13. Białek-Ławniczak P., Kuzemczak M., Torzyńska K., Janowska-Kulińska A., Miechowicz I., Kramer L., Moczko J., Siminiak T., Frequency domain HRV indices in stable coronary artery disease and their role in predicting long-term risk of urgent PCI – 5-year follow-up, <http://www.kardiologiapolska.pl/ptk2012/conf/16/P140.html>

14. Kuzemczak M., Białek-Ławniczak P., Torzyńska K., Janowska-Kulińska A., Miechowicz I., Kramer L., Moczko J., Siminiak T., The Allan factor – a novel HRV parameter predicting long – term risk of urgent PCI in patients with stable coronary artery disease and a history of myocardial infarction – 5- year follow-up, <http://www.kardiologiapolska.pl/ptk2012/conf/16/P149.html>

15. Białek-Ławniczak P., Kuzemczak M., Torzyńska K., Janowska-Kulińska A., Miechowicz I., Kramer L., Moczko J., Siminiak T., Oscillation index – a novel time domain HRV parameter and its role in predicting stroke in patients with stable coronary artery disease – 5-year follow-up, <http://www.kardiologiapolska.pl/ptk2012/conf/16/P027.html>

16. Kuzemczak M., Białek-Ławniczak P., Majewski M., Torzyńska K., Janowska-Kulińska A., Miechowicz I., Kramer L., Moczko J., Siminiak T., The analysis of correlation between severity of coronary atherosclerosis and novel HRV indices in patients with stable coronary artery disease, <http://www.kardiologiapolska.pl/ptk2012/conf/16/P145.html>

17. Schechtman V.L., Kluge K.A., Harper R.M., Time-domain system for assessing variation in heart rate, *Med. Biol. Comput.*, 1988; 26: 367 – 373

18. Saul J.P., Albrecht P., Berger R.D., Analysis of long-term heart rate variability: methods, *Computers in Cardiology*, 1987; IEEE Computer Society Press., Washington, 1988; 419 – 422

19. Malik M., Camm A.J., Components of heart rate variability – What they really mean and what we really measure, *Am. J. Cardiol.*, 1993; 72: 821 – 822
20. Kleiger R.E, Stein P.K. , Bigger J.T., Heart Rate Variability: Measurement and Clinical Utility, *Ann Noninvasive Electrocardiol.*, 2005;10(1):88-101
21. Heart rate variability; Standards of measurement, physiological interpretation, and clinical use. Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, *European Heart Journal*, 1996; 17: 354–381
22. Farrell T.G., Bashir Y., Cripps T., Malik M., i wsp., Risk stratification for arrhythmic events in postinfarction patients based on heart rate variability, ambulatory electrocardiographic variables and the signal-averaged electrocardiogram, *J. Am. Coll. Cardiol.*, 1991;18: 687
23. Odemuyiwa O., Malik M., Farrell T., Bashir Y., i wsp. Comparison of the predictive characteristics of heart rate variability index and left ventricular ejection fraction for all-cause mortality, arrhythmic events and sudden death after acute myocardial infarction, *Am. J. Cardiol.*, 1991;68: 434
24. Malik M., Camm J., Heart rate variability., *Clin. Cardiol.*, 1990;13: 570
25. Malik M., Farrell T., Cripps T., Camm A.J., Heart rate variability in relation to prognosis after myocardial infarction: selection of optimal processing techniques., *Eur. Heart J.*, 1989; 10:1060–74
26. Akselrod S., Gordon D., Ubel F.A., Shannon D.C., Barger A.C., Cohen R.J., Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat to beat cardiovascular control, *Science*, 1981; 213: 220 – 222
27. Sayers B. McA., Analysis of heart rate variability, *Ergonomics*, 1973;16(1): 17-32

28. Pomeranz B., Macaulay R.J.B., Caudill M.A., Kutz I., i wsp., Assessment of automatic function in humans by heart rate spectra analysis, *Am. J. Physiol.*, 1985; 248: H151 – H153
29. Pagani M., Lombardi F., Guzzetti S., Rimoldi O., i wsp., Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatio-vagal interaction in man and conscious dog, *Cir.Res.*, 1986; 59: 178 – 193
30. Malliani A., Pagani M., Lambardi F., Cerutti S., Cardiovascular neural regulation Explorer in the frequency domain, *Circulation*, 1991; 84: 482 – 492
31. Moczko J.A., Advanced methods of heart rate signals processing and their usefulness in diagnosis suport. I. Mathematical heart rate descriptors and virtual instrumentation, *Computational Methods in Science and Technology*, 2002; 8(2): 65 – 76
32. Masaoka S., Lev-Ran A., Hill L.R., Vakil G., Hon E.H., Heart rate variability in diabetes: relationship to age and duration of the disease, *Diabetes Care*, 1985; 8: 64 – 68
33. Montano N., Ruscone T.G., Porta A., Lombardi F., i wsp., Power spectrum analysis of heart rate variability to assess the changes in sympathovagal balance during graded orthostatic tilt, *Circulation*, 1994; 90: 1826 – 1831
34. Pagani M., Rimoldi O., Pizzinelli P., Furlan R., i wsp., Assessment of the neural control of the circulation during psychological stress, *J. Auton. Nerv. Syst.*, 1991; 35:33 – 41
35. Bigger J.T. Jr. Heart rate variability: frequency domain. W: Moss A.J., Stern S. red. *Noninvasive electrocardiology. Clinical aspects of Holter monitoring*, W.B. Saunders Comp., London 1996, 175
36. Akay M., *Time frequency and wavelets in biomedical signal processing*, IEEE Press., 1998
37. Qian S., Chen D., *Joint Time-Frequency Analysis*, PTR Prentice Hall, 1996

38. Moczko J.A., Kramer L., *Cyfrowe metody przetwarzania sygnałów biomedycznych*, Wydawnictwo Naukowe Poznań, 2001
39. Saul J.P., Albrecht P., Berger R.D., Cohen R.J., Analysis of long term heart rate variability: Methods, 1/f scaling and implications. *Comput Cardiol* 1987;14: 419–422
40. Bigger J.T., Steinman R.C., Rolnitzky L.M., Fleiss J.L., i wsp., Power law behaviour of RR-interval variability in healthy middle-aged persons, patients with recent myocardial infarction, and patients with heart transplants. *Circulation* 1996;93: 2142
41. Akay M., *Nonlinear biomedical Signac processing*, IEEE Press 2001; Vol. II, 6: 159 – 167
42. Makikallio T.H., Perkiomaki J.S., Huikuri H.V. Nonlinear dynamics of RR intervals. *Dynamic Electrocardiography*, Chapter 3, Futura 2004, pp. 22–30
43. Huikuri H.V., Mäkikallio T.H., Peng C-K, Goldberg A.L., i wsp., Fractal correlation properties of R-R interval dynamics and mortality in patients with depressed left ventricular function after an acute myocardial infarction. *Circulation* 2000;101: 47 – 53
44. Guzik P., Piskorski J., Krauze T., Schneider R., i wsp., Correlations between the Poicare Plot and conventional heart rate variability parameters assessed during paced breathing, *J. Physiol. Sci.*, 2007; 57: 63 – 71
45. Brennan M., Palaniswami M., Kamen P., Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Trans. Biomed. Eng.* 2001;48: 1342-7
46. Guzik P., Piskorski J., Krauze T., Bychowiec B., Wesseling KH., Schneider R., Girgus P., Wykrętowicz A., Wysocki H., Numerical descriptors of Poincare plots analysis of RR intervals are related to baroreflex sensitivity and hemodynamic parameters in health pe ople, *Folia Cardiolo.* 2005; 12 suppl.: 56-59

47. Toichi M., Sugiura T., Murai T., Sengoku A., A new method of assessing cardiac autonomic function and its comparison with spectra analysis and coefficient of variation of R-Rinterval, *J. Auton. Nerv. Syst.*, 1997; 62: 79-84
48. Bigger J.T., Fleiss J.L., Rolnitzky L.M., Steinman R.C., The ability of several short-term measures of RR variability to predict mortality after myocardial infarction, *Circulation*, 1993; 88: 927 – 934
49. Andrzejak R., Poręba R., Derkacz A., Poręba M., Skalik R., Turbulencja rytmu serca u chorych z pierwotnym nadciśnieniem tętniczym w zależności od typu geometrii i przerostu lewej komory serca (Heart rate turbulence in patients with essential hypertension and different patterns of left ventricular geometry and hypertrophy), *Nadciśnienie Tętnicze*, 2005; 9 (5): 356–363
50. Schmidt G., Malik M., Barthel P., Schneider R., i wsp., Heart rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction, *Lancet*, 1999; 353 – 1390
51. Davies L.C., Francis D.P., Ponikowski P., Piepoli M.F., Coats A.J., Relation of heart rate and blood pressure turbulence following premature ventricular complexes to baroreflex sensitivity in chronic congestive heart failure, *Am. J. Cardiol.* 2001;87: 737
52. Bauer A., Schmidt G. Heart rate turbulence in ischemic heart disease. *Dynamic Electrocardiography*, Chapter 23, Futura, 2004; 211–213
53. Herold G. i wsp. *Medycyna wewnętrzna. Repetytorium dla studentów medycyny i lekarzy.* Warszawa ; Wyd. Lek. PZWL; 2002
54. Bugajska J., Jędryka-Góral A., Konarska M., Występowanie czynników ryzyka choroby niedokrwiennej serca a praca zawodowa; *Bezpieczeństwo pracy*, 2006; 4: 14 – 17
55. Smith S.C., Greenland P., Grundy S.M., AHA Conference Proceedings. Prevention conference V: Beyond secondary prevention: Identifying the high-risk patient for primary prevention: executive summary. American Heart Association, *Circulation*, 2000; 111-115

56. Szostak W.B., Cybulska B. Otyłość wisceralna jako czynnik miażdżycy; „Medycyna po Dyplomie”; wyd. spec., 1996; 2-8
57. Cegła B., Federowicz G., Filanowicz M., Analiza czynników ryzyka choroby niedokrwiennej serca u pacjentów po 40 roku życia, *Zdrowie Publiczne*, 2011; 121(1): 52 – 57
58. Remme W.J. The sympathetic nervous system and ischemic heart disease, *Eur. Heart J.*, 1998; (suppl F): F62 – 71
59. Copie X., Lamaison D., Salvador M., Sadoul N., i wsp., Heart rate variability before ventricular arrhythmias in patients with coronary artery disease and an implantable cardioverter-defibrillator, *Ann. Noninvasive Electrocardiol.*, 2003; 8(3): 179 – 184
60. Casolo G., Stroder P., Signorini C., Calzolari F., i wsp., Heart rate variability during the acute phase of myocardial infarction, *Circulation*, 1992; 85(6): 2073-9
61. Flapan A.D., Wright R.A., Nolan J., Neilson J.M., Ewing D.J., Differing patterns of cardiac parasympathetic activity and their evolution in selected patients with a first myocardial infarction, *J. Am. Coll. Cardiol.*, 1993; 21: 926
62. Bigger J.T., Fleiss J.L., Rolnitzky L.M., Steinman R.C., Schneider W.J., Time course of recovery of heart period variability after myocardial infarction, *J. Am. Coll. Cardiol.*, 1991; 18: 1643 – 1649
63. Bigger J.T., Fleiss J.L., Steinman R.C., Rolnitzky L.M., i wsp., RR variability in healthy, middle-aged persons compared with patients with chronic coronary heart disease or recent acute myocardial infarction, *Circulation*, 1995; 91: 1936 – 1943
64. Bigger J.T.Jr, La Rovere M.T., Steinman R.C., Fleiss J.L., i wsp., Comparison of baroreflex sensitivity and heart period variability after myocardial infarction, *J. Am. Coll. Cardiol.*, 1989; 14: 1511

65. Bigger J.T., Fleiss J.L., Steinman R.C., Rolnitzky L.M., i wsp., Frequency domain measures of heart rate variability and mortality after myocardial infarction, *Circulation*, 1992; 85: 164 – 171
66. Singh N., Mironov D., Armstrong P.W., Ross A.M., Langer A., Heart rate variability assessment early after acute myocardial infarction, *Circulation*, 1996; 93: 1388-1395
67. Shehab A, Elnour AA, Struthers AD., Heart rate variability as an indicator of left ventricular systolic dysfunction, *Cardiovasc J Afr.* 2009 Sep-Oct;20(5):278-83
68. Huikuri H.V., Makikallio T., Airaksinen K.E.J., Mitrani R., i wsp., Measurement of heart rate variability: a clinical tool or research toy?, *J. Am. Coll. Cardiol.*, 1999; 34: 1878 – 1883
69. Tsuji H., Venditti F.J., Manders E.S., Evans J.C., i wsp., Reduced heart rate variability and mortality risk in an elderly cohort. The Framingham Heart Study, *Circulation*, 1994; 90: 878 – 883
70. Huikuri H.V., Jokinen V., Syvanne M., Nieminen M.S., i wsp., Heart rate variability and progression of coronary atherosclerosis, *Artheroscler. Thromb. Vasc. Biol.*, 1999; 19: 1979 – 1985
71. Dąbrowski A., Kramarz E., Wpływ aktywności życiowej na wartość rokowniczą zmienności rytmu serca u chorych po przebytych zawale serca, *Folia Cardiol.*, 2003;10: 45
72. Faber T.S., Staunton A., Hnatkova K., Camm A.J., Malik M., Stepwise strategy of using short- and long-term heart rate variability for risk stratification after myocardial infarction, *Pacing Clin Electrophysiol.*, 1996; 19: 1845 – 1851
73. Fei L., Copie X., Malik M., Camm A.J., Short- and long-term assessment of heart rate variability for risk stratification after acute myocardial infarction, *Am. J. Cardiol.*, 1996; 77: 681 – 684

74. Neaton O.D., Grimm R.H., Prineas R.J., Stamler J., i wsp., Treatment of Mild Hypertension Study. Final results, JAMA, 1993; 270: 713–724
75. Bellwon J., Siebert J., Jagodziński P., Rykaczewski A., i wsp. Kumulacja czynników ryzyka choroby wieńcowej w nadciśnieniu tętniczym, Nadciśnienie Tętnicze, 1998; 2: 21–26
76. British Hypertension Society guidelines for hypertension management 2004 (BHS-IV): summary, BMJ. 2004; 328(7440): 634–640
77. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure
78. Piotrowicz R., Stolarz K., Zmienność rytmu serca w nadciśnieniu tętniczym. Część II: Zmienność rytmu serca i jej profil dobowy u pacjentów z nadciśnieniem tętniczym, Nadciśnienie Tętnicze, 1999; 3(4): 257 – 264
79. Tsuji H., Larson M.G., Venditti F.J., Manders E.S., i wsp. Impact of reduced heart rate variability on risk for cardiac events: the Framingham Heart Study. Circulation 1996; 94: 2850–2855
80. Dekker J.M., Crow R.S., Folsom A.R., Hannan P.J., i wsp. Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes. The ARIC Study. Circulation 2000; 102: 1239–1244
81. Piotrowicz R., Stolarz K., Zmienność rytmu serca w nadciśnieniu tętniczym Część V. Znaczenie prognostyczne zmienności rytmu serca, Nadciśnienie Tętnicze, 2004; 8(1): 55 – 59
82. Mancia G., Ferrari A., Gregorini L., Parati G., i wsp., Blood pressure and heart rate variability in normotensive and hypertensive human beings, Circ. Res., 1983; 53: 96 – 104
83. Radaelli A., Bernardi L., Valle F., Leuzzi S., i wsp., Cardiovascular autonomic modulation in essential hypertension. Effect of tilting, Hypertension, 1994; 24: 556 – 563

84. Huikuri H.K., Ylitalo A., Pikkujämsä S.M., Ikäheimo M.J., i wsp., Heart rate variability in systemic hypertension. *Am. J. Cardiol.*, 1996; 77: 1073–1077
85. Langewitz W., Rüddel H., Schächinger H.: Reduced parasympathetic cardiac control in patients with hypertension at rest and under mental stress, *Am. Heart J.*, 1994; 127: 122–128
86. Dassi S., Balsamà M., Guzzetti S., Ponti G.B., i wsp., Twenty-four hor power spectral analysis of heart rate variability and of arteria, *J. Hypertens.*, 1991; 9 (supl. 6): S72–S73
87. Guzzetti S., Piccaluga E., Casati R., Cerutti S., i wsp., Sympathetic predominance in essential hypertension: a study employing spectra analysis of heart rate variability, *J. Hypertens.*, 1988; 6: 711–717
88. Yo Y., Nagano M., Nagano N., Iiyama K., i wsp., Effects of age and hypertension on autonomic nervous regulation during passive head-up tilt., *Hypertension*, 1994; 23 (supl. I): I82–I86
89. Takalo R., Korhonen I., Turjanmaa V., Majahalme S., Tuomisto M., Uusitalo A., Short-term variability of blood pressure and heart rate in borderline and mildly hypertensive subjects, *Hypertension*, 1994; 23: 18 – 24
90. Furlan R., Guzzetti S., Crivellaro W., Dassi S., i wsp., Continuous 24-hour assessment of the neural regulation of systemic arteria pressure and RR variabilities in ambulant subjects, *Circulation*, 1990; 81: 537 – 547
91. Piccirillo G., Munizzi M.R., Fimorgani F.L., Marigliano V., Heart rate variability in hypertensive subjects. *Int. J. Cardiol.*, 1996; 53: 291–298
92. Guzzetti S., Piccaluga E., Casati R., Cerutti S., i wsp., Sympathetic predominance in essential hypertension: a study employing spectra analysis of heart rate variability. *J. Hypertens.* 1988; 6:, 711–717

93. Takalo R., Korhonen I., Turjanmaa V., Majahalme S., Tuomisto M., Uusitalo A., Short-term variability of blood pressure and heart rate in borderline and mildly hypertensive subjects, *Hypertension*, 1994; 23: 18–24
94. Korzeniowska K., Jabłeczka A., Cukrzyca (Część I), *Farm.Współ.*, 2008; 1: 231-235
95. Wheeler T., Watkins P.J., Cardiac denervation in diabetes, *BMJ*, 1973; 4: 584 – 586
96. Ewing D.J., Campbell I.W., Burt A.A., Clarke B.F., Vascular reflexes in diabetic autonomic neuropathy, *Lancet*, 1973; 2: 1354 – 1356
97. Liao D., Cai J., Brancati F., Folsom A., i wsp., Association of vagal tone with serum insulin, glucose and diabetes mellitus. The ARIC Study, *Diabetes Res. Clin. Pract.*, 1995; 30: 211 – 221
98. Ziegler D., Diabetic cardiovascular autonomic neuropathy: prognosis, diagnosis and treatment, *Diabetes Metab. Rev.*, 1994; 10: 339 – 383
99. Defonzo R.A., Ferramini E., Insulin resistance: a multiple faced syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia and atherosclerotic cardiovascular disease, *Diabetes Care*, 1991; 14: 173 – 194
100. Folsom A.R., Szklo M., Stevens J., Liao F., i wsp., A prospective study of coronary heart disease in relation to fasting insulin, glucose and diabetes. The atherosclerosis risk in communities (ARIC) study, *Diabetes Care*, 1997; 20: 935 – 942
101. Majkowska L., Wpływ leczenia nadciśnienia tętniczego u chorych na cukrzycę typu 2 na rozwój powikłań o typie makroangiopatii — wyniki dużych badań interwencyjnych, *nadciśnienie tętnicze*, 2002, 6(3): 205 – 215
102. Kataoka M., Ito C., Sasaki H., Yamane K., Kohno N., Low heart rate variability is a risk factor for sudden cardiac death in a type 2 diabetes, *Diabetes Res. Clin. Pract.*, 2004; 64(1): 51 – 58

103. Liao D., Sloan R.P., Cascio W.E., Folsom A.R., i wsp., Multiple metabolic syndrom eis associated with heart rate variability. The atherosclerosis RSC in communities study, *Diabetes Care*, 1998; 21 (21): 2116 – 2122
104. Beck M.O., Silveiro S.P., Friedman R., Clausell N., Gross J.L., Asymptomatic coronary artery disease is associated with cardiac autonomic neuropaty and diabetic nephropathy in type 2 diabetic patients, *Diabetes Care*, 1999; 22: 1745 – 1747
105. Kempler P., Váradi A., Szalay F., Tamás G., Autonomic neuropaty and corrected QT interwal prolongation: there is a relationship, *Diabetes Care*, 1994; 17: 454 – 456
106. Khandoker A.H., Jelinek H.F., Palaniswami M., Identifying diabetic patients with cardiac autonomic neuropaty by heart rate complexity analysis; *BioMedical Engineering OnLine*, 2009, 8:3
107. Yeh S.Y., Forsythe A., Hon E.H., Quantification of fetal heart rate beat-to-beat interwal differences, *Obstet. Gynecol.*, 1973; 41(3): 355 – 363
108. de Haan J., van Bommel J.H., Stolte L.A.M., Janssens J., i wsp., Quantitative evaluation of fetal heart rate patterns III. Beat-to-beat arrhythmia, *Europ. J. Obstet. Gynec.*, 1971; 4: 137 – 146
109. Organ L.W., Hawrylyshyn P.A., Goodwin J.W., Milligan J.E., Bernstein A., Quantitative indices of short- and long-term heart rate variability, *Am. J. Obstet. Gynec.*, 1978; 130(1): 20 – 27
110. van Geijn H.P., Analysis of heart rate and beat-to-beat variability. Interval difference index, *Am. J. Obstet. Gynec.*, 1980; 138(3): 246 – 252
111. Huey J.R., Paul R.H., Hadjiev A.A., Jilek J., Hon E.H., Fetal heart rate variability. An approach to automatem assessment, *Am. J. Obstet. Gynec.*, 1979; 134(6): 691 – 695
112. Dalton K.J., Dawes G.S., Patrick J.E., Diurnal respiratory and Rother rhythms of fetal heart rate in lambs, *Am. J. Obstet. Gynec.*, 1977; 127(4): 414 – 424

113. Zugaib M., Forsythe A.B., Nuwayhid B., Lieb S.M., i wsp., Mechanisms of beat-to-beat variability in the heart rate of the neonatal lamb. I Influence of the autonomic nervous system, *Am. J. Obstet. Gynec.*, 1980; 138(4): 444 – 452
114. Pomeranz B., Macaulay R.J.B., Caudill M.A., Kutz I., i wsp., Assessment of autonomic function in humans by heart rate spectra analysis. *Am. J. Physiol.*, 1985; 248: H151 - H153
115. Piotrowicz R., Stolarz K.: Zmienność rytmu serca w nadciśnieniu tętniczym. część I. Wprowadzenie. *Nadciśnienie Tętnicze*, 1999; 3: 194 – 199
116. Olansen J.B., Rosow E., *Virtual Bio-Instrumentation. Biomedical, clinical and Healthcare applications in LabVIEW*, Prentice Hall PTR, 2002
117. Chugani M.I., Samant A.R., Cerna M., *LabVIEW Signal Processing*, Prentice Hall PTR, New York, 1998
118. Horoba K., Graczyk S., Jeżewski J., Wróbel J., *Biomedical Signal acquisition and analysis using LabVIEW tools: an application to abdominal electrohysterography*, *Archives of Perinatal Medicine*, 1997; 3: 65 – 73
119. Horoba K., Wróbel J., Jeżewski J., Gacek A., i wsp., *Acquisition and processing of cardiotocographic data in LabVIEW environment*, *Proc. of the 4th International Conference on Computers in Medicine*, Zakopane, 1997; 1: 141 – 145
120. Cochran W. G., *The chi-square goodness-of-fit test*, *Annals of Mathematical Statistics*, 1952, 23(3), 15-345
121. Fisher R.A., *Statistical methods for research workers (5th edition)*, Oliver and Boyd, Edinburgh, 1934
122. Fisher R.A., *The logic of inductive inference*, *Journal of the Royal Statistical Society Series A*, 1935; 98: 39-54

123. Verhulst P.F., Notice sur la loi que la population poursuit dans son accroissement, *Correspondance Mathematique et Physique*, 10, 113 – 121
124. Berkson J., Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, 1944; 39, 357-365
125. Stanisław A., Przystępny kurs statystyki z wykorzystaniem STATISTICA.PL na przykładach z medycyny, tom II, Wydawnictwo StatSoft, Karków, 2000
126. Ferguson G.A., Takane Y., *Analiza statystyczna w psychologii i pedagogice*, PWN, Warszawa, 1999
127. Hosmer D.W., Lemeshow S., *Applied logistic regression*, wyd.2, Wiley & Sons, New York, 2000
128. Stanisław A., *Regresja logistyczna*, *Medycyna Praktyczna*, 2001, (10)
129. Kleinbaum D.G., Klein M., *Logistic regression – a self-learning text*, wyd 2, Springer, New York, 2002
130. Danieluk B., *Zastosowanie regresji logistycznej w badaniach eksperymentalnych*, *Psychologia społeczna*, tom 5, 2010; 2-3 (14):199-216
131. Breiman L., Friedman J.H., Olshen R.A., Stone C.J., *Classification and regression trees*, Wyd. Chapman & Hall, Londyn, 1993
132. Loh W.Y, Shih Y.S. 1997. Split selection methods for classification trees. *Statistica Sinica*, 7. s. 815-840.
133. Koronacki J., Ćwik J., *Statystyczne systemy uczące się*, Wyd. Naukowo-Techniczne, Warszawa, 2005
134. Internetowy Podręcznik Statystyki; <http://www.statsoft.pl/textbook>

135. Tufféry S., Data mining and statistics for decision making, Wiley, 2011
136. Lewek J., Wranicz J.K., Guzik P., Chudzik M., i wsp., Clinical and electrocardiographic covariates of deceleration capacity in patients with ST-segment elevation myocardial infarction, *Cardiology J.*; 2009; 16(6): 528-534
137. Dos Santos L., Barroso J.J., Macau E.E., de Godoy M.F., Application of an autonomic adaptive filter for Heart Rate Variability analysis, *Med. Eng. Phys.*, 2013;35(12):1778-85
138. Tarvainen M.P., Niskanen J.P., Lipponen J.A., Ranta-Aho P.O. i wsp., Kubios HRV – Heart rate variability analysis software, *Comput. Methods Programs Biomed.*, 2014;113(1):210-20
139. Huikuri H.V., Stein P.K., Clinical application of heart rate variability after acute myocardial infarction, *Front Physiol*, 2012;3:41
140. Chiang J.K., Fu C.H., Kuo T.B., Koo M., Association between frequency domain heart rate variability and unplanned readmission to hospital in geriatric patients, *BMC Public Health*, 2011; 11:137
141. Chrysohoou Ch., Skoumas J., Oikonomou E., Tsiachris D., i wsp., Aortic artery distensibility shows inverse correlation with heart rate variability in elderly hypertensive individuals. The IKARIA Study., *Heart and Vessels*, 2013;28: 467-472
142. Turcott R.G., Teich M.C., Fractal character of the electrocardiogram: distinguishing heart-failure and normal patients, *Ann. Biomed.Eng.*,1996; 24 (2): 269-293
143. Molgaard H., Hermansen K., Bjerregaard P., Spectral components of short-term RR interval variability in healthy subjects and effects of risk factors, *Eur. Heart J.*, 1994;15:1174
144. Bonnemeier H., Richardt G., Potratz J., et al., Circadian profile of cardiac autonomic nervous modulation in health subjects: differing effects of aging and gender on heart rate variability, *J. Cardiovasc Electrophysiol*, 2003;14:791-799

145. Umotani K., Singer D.H., McCraty R., Twenty-four hour time domain heart rate variability and heart rate: relations to age and gender over nine decades, *J. Am. Coll. Cardiol*, 1998; 31:593
146. Voss A., Schroeder R., Fischer C., Heitmann A., i wsp. Influence of age and gender on complexity measures for short term heart rate variability in healthy subjects, *Conf Proc IEEE Eng Med Biol Soc.* 2013;2013:5574-7
147. Stein P.K., Kleiger R.E., Rottman J.N., Differing effects of age on heart rate variability in men and women, *Am. J. Cardiol*, 1997;80:302-305
148. Huikuri H.V., Pikkujämsä S.M., Airaksinen K.E., i wsp., Sex-related differences in autonomic modulation of heart rate in middle-aged subjects, *Circulation*, 1996; 94:122-125
149. Antelmi J., de Paula R.S., Shinzato A.R., Peres C.A., i wsp., Influence of age, gender, body mass index and functional capacity on heart rate variability in a cohort and subjects without heart disease, *Am J. Cardiol* 2004;1:93(3): 381-385
150. Poirier P., Giles T., Bray G., Hong Y., i wsp., Obesity and Cardiovascular Disease: Pathophysiology, Evaluation and Effect of Weight Loss, *Circulation* 2006;113:898-918
151. Hirsch J., Leibel R.L., Macintosh R., Aguirre A., i wsp., Heart rate variability as a measure of autonomic function during weight change in humans, *Am J. Physiol.*, 1991;216:1418-1423
152. Gerritsen J., Dekker J.M., Ten Voorde B.J., Bertelsmann F.W., i wsp., Glucose tolerance and other determinants of cardiovascular autonomic function: the Horn Study, *Diabetologia* 2000;43:561-570
153. Singh J.P., Larson M.G., O'Donnell C.J., Wilson P.F., i wsp., Association of hyperglycemia with reduced heart rate variability: the Framingham Heart Study, *Am. J. Cardiol.*, 2000, 86: 309-312

154. Liao D.P., Cai J.W., Brancati F.L., Folsom A., i wsp., Association of vagal tone with serum insulin glucose and diabetes mellitus; the ARIC Study, *Diabetes Res. Clin. Prac.* 1995; 30:211-221
155. Muscelli E., Emalin M., Notali A., Pratali L., i wsp., Autonomic et hemodynamic responses to insulin in lean and obese humans, *J. Clin. Endocrinol. Metab.*, 1998; 83(6):2084-2090
156. Baumert M., Sacre J.W., Heart rate complexity and cardiac sympathetic dysinnervation in patients with type 2 diabetes mellitus, *Conf Proc IEEE Eng Med Biol Soc.* 2013;2013:5570-3
157. Roy B., Ghatak S., Nonlinear methods to assess changes in heart rate variability in type 2 diabetic patients, *Arq Bras Cardiol.* 2013;101(4):317-27
158. Turker Y., Aslantas Y., Aydin Y., Demirin H., i wsp., Heart rate variability and heart rate recovery in patients with type 1 diabetes mellitus, *Acta kardiol.*, 2013; 68(2):145-50
159. Osterheus H., Kochs M., Hombach V., Time-dependent changes of heart rate variability after percutaneous trans luminal angioplasty, *A. Heart J.* 1998; 135: 755-761
160. Oliveira N.L., Ribeiro F., Alves A.J., Teixeira M., Heart rate variability in myocardial infarction patients: effect of exercise training, *Rev Port Cardiol.* 2013 Sep;32(9):687-700
161. Dąbrowska B., Dąbrowski A., Piotrowicz R., „Elektrokardiografia holterowska”, ViaMedica, Gdańsk 2004; 189-190
162. Pawlak-Buś K., Zmienność rytmu serca u pacjentów ze stabilną chorobą niedokrwinną. Nowe modele interpretacji zjawiska w ocenie równowagi autonomicznej, Praca doktorska, Poznań, 2006

163. Pal G.K., Adithan C, Dutta TK, Pal P. i wsp., Association of hypertension status and cardiovascular risks with sympathovagal imbalance in first degree relatives of type 2 diabetics, *J Diabetes Investig.*, 2014; 5(4):449-455
164. Jaccard J., *Interaction Effects in Logistic Regression*, Sage University Papers, Series on Quantitative Applications in the Social Sciences, 07-135, Thousand Oaks, 2001
165. Jackowska B., Efekty interakcji między zmiennymi objaśniającymi w modelu logitowym w analizie zróżnicowania ryzyka zgonu, *Przegląd Statystyczny*, 2011; 58 (1-2): 24-41
166. Więckowska B., *Metody statystyczne w poszukiwaniu czynników ryzyka występowania wrodzonych wad rozwojowych na podstawie danych z Polskiego Rejestru Wrodzonych Wad Rozwojowych*, rozprawa doktorska, Poznań, 2011
167. Sowińska A., *Ocena przydatności wskaźników wagowo – wzrostowych i analiza wybranych modeli statystycznych zastosowanych do wyznaczania czynników ryzyka występowania nieprawidłowego ciśnienia tętniczego krwi u dzieci*, rozprawa doktorska, Poznań, 2011

Załącznik 1 Tabele przedstawiające wystąpienie istotnych statystycznie różnic w wartościach parametrów pomiędzy osobami z grupy badanej i kontrolnej

1. pacjenci po przebytych zawale mięśnia sercowego

Tabela A 1 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród pacjentów po przebytych zawale mięśnia sercowego i pacjentów z grupy kontrolnej (test U Manna-Whitneya)

cecha	Osoby bez zawału, nadciśnienia i cukrzycy (N=62)				Osoby, które przeszły zawał (N=84)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
wiek	56,1 ± 10,2	55,0	35,0	88,0	59,2 ± 9,7	57,0	35,0	82,0	0,039
BMI	25,8 ± 2,8	26,0	20,0	34,4	27,6 ± 4,3	27,0	18,7	38,3	0,011
de Haan STI	0,020 ± 0,010	0,020	0,050	0,040	0,010 ± 0,010	0,010	0,003	0,050	0,007
de Haan LTI	54,9 ± 28,2	47,7	16,5	159,4	43,9 ± 27,5	35,3	8,3	163,2	0,005
Yeh DI	0,020 ± 0,010	0,010	0,004	0,070	0,010 ± 0,010	0,010	0,002	0,050	0,023
Yeh II	0,040 ± 0,020	0,030	0,010	0,150	0,030 ± 0,010	0,030	0,008	0,070	0,030
van Geijn ID	3,82 ± 1,80	3,50	1,25	9,66	3,00 ± 1,40	2,77	1,07	7,17	0,002
Huey LTV	89,00 ± 45,00	80,60	22,10	226,95	71,50 ± 38,80	64,50	18,50	204,58	0,014
Dalton MABB	11,00 ± 6,90	9,95	2,25	37,37	8,18 ± 5,30	7,17	1,09	26,74	0,005
Dalton SD	35,50 ± 22,60	31,30	9,70	149,50	27,55 ± 15,30	23,85	5,66	92,04	0,017
Zugaib STV	0,007 ± 0,004	0,006	0,002	0,023	0,005±0,003	0,004	0,001	0,020	0,009
Zugaib LTV	0,030 ± 0,010	0,030	0,008	0,070	0,020 ± 0,010	0,020	0,007	0,060	0,009
SDNN	18,80 ± 13,80	16,00	4,81	73,55	15,20 ± 11,20	12,00	2,85	73,72	0,045

2. pacjenci po przebytych zawale mięśnia sercowego chorujący na nadciśnienie

Tabela A 2 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie i osób z grupy kontrolnej (test U Manna-Whitneya)

cecha	Osoby bez zawału, nadciśnienia i cukrzycy (N=62)				Osoby, które chorują na zawał i nadciśnienie (N=33)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
wiek	56,1 ± 10,2	55,0	35,0	88,0	60,5 ± 10,4	59,0	35,0	82,0	0,028
de Haan STI	0,020 ± 0,010	0,020	0,050	0,040	0,010 ± 0,010	0,010	0,004	0,030	0,048
de Haan LTI	54,9 ± 28,2	47,7	16,5	159,4	41,7 ± 25,5	32,5	14,0	112,1	0,007
van Geijn ID	3,82 ± 1,80	3,50	1,25	9,66	3,00 ± 1,40	2,77	1,12	7,17	0,018
Dalton MABB	11,00 ± 6,90	9,95	2,25	37,37	8,13 ± 4,00	8,14	2,26	18,60	0,044
Zugaib LTV	0,030 ± 0,010	0,030	0,008	0,070	0,020 ± 0,010	0,020	0,005	0,010	0,039

3. pacjenci po przebytych zawale mięśnia sercowego chorujący na nadciśnienie tętnicze i cukrzycę

Tabela A 3 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie i cukrzycę oraz osób z grupy kontrolnej (test U Manna-Whitneya)

cecha	Osoby bez zawału, nadciśnienia i cukrzycy (N=62)				Osoby, które chorują na zawał, nadciśnienie i cukrzycę (N=18)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
wiek	56,1 ± 10,2	55,0	35,0	88,0	61,3 ± 7,8	62,0	49,0	74,0	0,022
BMI	25,8 ± 2,8	26,0	20,0	34,4	29,8 ± 4,3	28,1	24,4	37,0	0,001
de Haan STI	0,020 ± 0,010	0,020	0,050	0,040	0,010 ± 0,010	0,008	0,003	0,050	0,002
de Haan LTI	54,9 ± 28,2	47,7	16,5	159,4	39,8± 25,4	35,3	8,3	96,6	0,032
Yeh DI	0,020 ± 0,010	0,010	0,004	0,070	0,010 ± 0,010	0,006	0,002	0,040	0,008
Yeh II	0,040 ± 0,020	0,030	0,010	0,150	0,030 ± 0,010	0,020	0,009	0,060	0,015
van Geijn ID	3,82 ± 1,80	3,50	1,25	9,66	2,70± 1,40	2,50	1,07	6,10	0,011
Huey STV	99,1 ± 68,0	81,0	22,3	340,2	70,0 ± 40,6	52,5	27,1	158,0	0,047
Huey LTV	89,00 ± 45,00	80,60	22,10	226,95	61,90 ± 36,80	54,20	27,20	148,40	0,010
Dalton MABB	11,00 ± 6,90	9,95	2,25	37,37	6,45 ± 6,10	3,80	1,10	26,70	0,001
Dalton SD	35,50 ± 22,60	31,30	9,70	149,50	23,00 ± 14,50	17,00	5,66	58,30	0,007
Zugaib STV	0,007 ± 0,004	0,006	0,002	0,023	0,004 ± 0,003	0,003	0,001	0,010	0,002
Zugaib LTV	0,030 ± 0,010	0,030	0,008	0,070	0,020 ± 0,010	0,020	0,007	0,050	0,020
OSC	12,2 ± 8,4	9,9	2,5	46,9	9,7 ± 8,3	6,7	3,4	39,6	0,043
SDNN	18,80 ± 13,80	16,00	4,81	73,55	11,40 ± 7,20	8,70	2,80	28,90	0,006
RMSSD	51,9 ± 19,0	45,6	28,7	119,0	40,3 ± 11,4	40,0	23,4	67,5	0,004
wsk. Allana	0,11 ± 0,02	0,11	0,06	0,14	0,12 ± 0,01	0,12	0,09	0,13	0,023

4. pacjenci chorujący na nadciśnienie tętnicze i cukrzycę

Tabela A 4 Wystąpienie istotnych statystycznie różnic w wartościach parametrów wśród pacjentów chorujących na nadciśnienie i cukrzycę oraz osób z grupy kontrolnej (test U Manna-Whitneya)

Cecha	Osoby bez zawału, nadciśnienia i cukrzycy (N=62)				Osoby, które chorują na nadciśnienie i cukrzycę (N=42)				p
	średnia±SD	mediana	min	maks	średnia±SD	mediana	min	maks	
wiek	56,1 ± 10,2	55,0	35,0	88,0	60,6 ± 9,4	60,5	47,0	81,0	0,020
BMI	25,8 ± 2,8	26,0	20,0	34,4	30,0 ± 4,4	29,7	19,7	37,4	<0,001
de Haan STI	0,020 ± 0,010	0,020	0,050	0,040	0,010 ± 0,010	0,010	0,003	0,050	<0,001
de Haan LTI	54,9 ± 28,2	47,7	16,5	159,4	36,1 ± 21,0	32,8	5,4	102,4	<0,001
Yeh DI	0,020 ± 0,010	0,010	0,004	0,070	0,010 ± 0,010	0,007	0,002	0,050	<0,001
Yeh II	0,040 ± 0,020	0,030	0,010	0,150	0,030 ± 0,020	0,022	0,007	0,110	<0,001
Organ BAND	4,90 ± 2,60	4,40	1,13	13,17	3,60 ± 1,80	3,00	0,74	8,90	0,007
van Geijn ID	3,82 ± 1,80	3,50	1,25	9,66	2,90 ± 1,70	2,30	0,86	8,80	0,001
Huey STV	99,1 ± 68,0	81,0	22,3	340,2	75,1 ± 42,9	60,6	12,8	193,1	0,045
Huey LTV	89,00 ± 45,00	80,60	22,10	226,95	69,10 ± 39,30	57,80	13,70	170,40	0,017
Dalton MABB	11,00 ± 6,90	9,95	2,25	37,37	6,40 ± 4,40	5,30	1,03	19,80	<0,001
Dalton SD	35,50 ± 22,60	31,30	9,70	149,50	22,50 ± 15,50	19,60	4,40	84,20	<0,001
Zugaib STV	0,007 ± 0,004	0,006	0,002	0,023	0,004 ± 0,003	0,003	0,001	0,013	<0,001
Zugaib LTV	0,030 ± 0,010	0,030	0,008	0,070	0,020 ± 0,010	0,020	0,005	0,050	<0,001
OSC	12,2 ± 8,4	9,9	2,5	46,9	8,8 ± 6,5	6,8	1,6	39,7	0,005
SDNN	18,80 ± 13,80	16,00	4,81	73,55	13,80 ± 11,60	10,60	2,17	53,60	0,003
RMSSD	51,9 ± 19,0	45,6	28,7	119,0	43,8 ± 13,2	41,3	22,3	87,4	0,025
wsk. Allana	0,11 ± 0,02	0,11	0,06	0,14	0,12 ± 0,02	0,12	0,07	0,15	0,020

Załącznik 2 Budowa analizowanych modeli logistycznych oraz tabele korelacji między zmiennymi uwzględnionymi w modelu

1. pacjenci po przebytych zawale mięśnia sercowego

zmienne BMI, de Haana STI

N=146	Model: Regr. logistyczna (logit) N zer: 62 jedynek: 84 (OBLICZENIA KONCOWE po 2 korekcie styczen2006) Zmn. zal.: zawał_zdrowe Strata: Największe prawd. bł.średnkw.skal. Całkowita strata: 92,387564894 Chi2(2)=14,296 p=,00079		
	Stała B0	BMI	pr1_de Hann STI
Ocena	-2,627045	0,1372494	-44,62405
Błąd standard.	1,351346	0,04948963	19,78166
t(143)	-1,94402	2,773295	-2,25583
poziom p	0,05385628	0,0062895	0,02559883
-95%CL	-5,298241	0,03942359	-83,72631
+95%CL	0,04415095	0,2350751	-5,521802
Chi-kwadrat Walda	3,779215	7,691167	5,088768
poziom p	0,05190174	0,00555238	0,02408802
Iloraz szans z.jedn.	0,07229178	1,147114	4,1688750E-20
-95%CL	0,00500038	1,040211	4,3463640E-37
+95%CL	1,04514	1,265004	0,003998633
Iloraz szans zakr.		14,84548	0,122237
-95%CL		2,170349	0,01938009
+95%CL		101,5451	0,7709923

Tabela A 5 Wartość współczynnika R Spearmana

Para zmiennych	Korelacja porządku rang Spearmana Oznaczone wsp. korelacji są istotne z p <0,05			
	N ważnych	R Spearman	t(N-2)	p
BMI & de Haan STI	146	-0,090	-1,08563	0,279

zmienne BMI, de Haana LTI

Model: Regr. logistyczna (logit) N zer: 62 jedynek: 84 (OBLICZENIA KONCOWE po 2 korekcie styczen201			
Zmn. zal.: zawał_zdrowe Strata: Największe prawd. bł.średnkw.skal.			
Całkowita strata: 91,851659582 Chi2(2)=15,368 p=,00046			
N=146	Stała B0	BMI	pr1_de Hann LTI
Ocena	-2,947969	0,150958	-0,01565931
Błąd standard.	1,357522	0,05106754	0,006402329
t(143)	-2,171582	2,956047	-2,445878
poziom p	0,03153629	0,00364607	0,0156642
-95%CL	-5,631371	0,05001322	-0,02831475
+95%CL	-0,264567	0,2519028	-0,00300388
Chi-kwadrat Walda	4,715768	8,738211	5,982317
poziom p	0,02989453	0,00311837	0,01445535
Iloraz szans z.jedn.	0,0524461	1,162948	0,9844627
-95%CL	0,00358366	1,051285	0,9720824
+95%CL	0,7675382	1,286471	0,9970006
Iloraz szans zakr.		19,43636	0,08841445
-95%CL		2,672546	0,01244892
+95%CL		141,3529	0,6279351

Tabela A 6 Wartość współczynnika R Spearmana

Para zmiennych	Korelacja porządku rang Spearmana Oznaczone wsp. korelacji są istotne z p <0,05			
	N ważnych	R Spearman	t(N-2)	p
BMI & de Haan LTI	146	0,054	0,648150	0,518

zmienne BMI, Yeh II

Model: Regr. logistyczna (logit) N zer: 62 jedynek: 84 (OBLICZENIA KONCOWE po 2 korekcie styczen201			
Zmn. zal.: zawał_zdrowe Strata: Największe prawd. bł.średnkw.skala.			
Całkowita strata: 91,745665424 Chi2(2)=15,580 p=,00041			
N=146	Stała B0	BMI	pr1_Yeh II
Ocena	-2,690416	0,1454491	-26,40595
Błąd standard.	1,359503	0,05051968	11,05715
t(143)	-1,97897	2,879059	-2,388134
poziom p	0,04974009	0,00460176	0,01824014
-95%CL	-5,377735	0,0455873	-48,26253
+95%CL	-0,00309663	0,245311	-4,549375
Chi-kwadrat Walda	3,916322	8,288983	5,703185
poziom p	0,04782794	0,00399127	0,01693992
Iloraz szans z.jedn.	0,06785271	1,156559	3,40440300E-12
-95%CL	0,004618269	1,046642	1,09610000E-21
+95%CL	0,9969081	1,278019	0,01057381
Iloraz szans zakr.		17,44174	0,0219651
-95%CL		2,449878	0,0009314559
+95%CL		124,1753	0,5179691

Tabela A 7 Wartość współczynnika R Spearmana

Para zmiennych	Korelacja porządku rang Spearmana Oznaczone wsp. korelacji są istotne z p <0,05			
	N ważnych	R Spearman	t(N-2)	p
BMI & Yeh II	146	0,030	0,363073	0,717

zmienne BMI, van Geijn ID

Model: Regr. logistyczna (logit) N zer: 62 jedynek: 84 (OBLICZENIA KONCOWE po 2 korekcje styczen2011)			
Zmn. zal.: zawał_zdrowe Strata: Największe prawd. bł.średnkw.skal.			
Całkowita strata: 90,146949241 Chi2(2)=18,777 p=,00008			
N=146	Stała B0	BMI	pr1_van Geijn ID
Ocena	-2,349023	0,142904	-0,3449053
Błąd standard.	1,36284	0,05026472	0,1158813
t(143)	-1,723623	2,843028	-2,976366
poziom p	0,08693715	0,00512352	0,003426237
-95%CL	-5,042938	0,0435461	-0,573967
+95%CL	0,3448924	0,2422618	-0,1158435
Chi-kwadrat Walda	2,970877	8,082806	8,858752
poziom p	0,08478554	0,00447155	0,00291905
Iloraz szans z.jedn.	0,09546243	1,153619	0,7082875
-95%CL	0,00645476	1,044508	0,5632864
+95%CL	1,411838	1,274128	0,8906146
Iloraz szans zakr.		16,59065	0,05160609
-95%CL		2,353533	0,007207217
+95%CL		116,9518	0,3695168

Tabela A 8 Wartość współczynnika R Spearmana

Para zmiennych	Korelacja porządku rang Spearmana Oznaczone wsp. korelacji są istotne z p <0,05			
	N ważnych	R Spearman	t(N-2)	p
BMI & van Geijn ID	146	-0,102	-1,23397	0,219

2. pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie

zmienne BMI, de Haana STI, Yeh DI, Organ BAND, van Geijn ID, Zugaib LTV, wiek

N=95	Model: Regr. logistyczna (logit) N zer: 62 jedynek: 33 (dane do doktoratu) Zmn. zal.: zawał i nadciśnienie/zdrowe Strata: Największe prawd. bł.średnkw.skala. Całkowita strata: 49,465246532 Chi2(7)=23,772 p=,00125							
	Stała B0	BMI	pr1_de Hann STI	pr1_Yeh DI	pr1_Organ BAND	pr1_van Geijn ID	pr1_Zugaib LTV	wiek
Ocena	-7,921436	0,1674277	1,833706E+02	-124,1466	1,394372E+00	-0,6789917	-239,2032	0,05905789
Błąd standard.	2,98742	0,0712747	8,167193E+01	57,44232	4,584466E-01	0,3378199	81,11778	0,02805397
t(87)	-2,651598	2,349047	2,245210E+00	-2,161239	3,041515E+00	-2,009922	-2,948838	2,105153
p	0,009518768	0,021087	2,728980E-02	0,03342419	3,110762E-03	0,04753761	0,004095657	0,03816121
-95%CL	-13,85926	0,0257614	2,103880E+01	-238,3194	4,831601E-01	-1,350445	-400,4336	0,00329759
+95%CL	-1,983616	0,309094	3,457024E+02	-9,973737	2,305584E+00	-0,007538086	-77,97284	0,1148182
Chi-kwadrat Walda	7,030973	5,51802	5,040967E+00	4,670953	9,250815E+00	4,039787	8,695646	4,431669
p	0,008015128	0,0188276	2,476147E-02	0,03068429	2,355811E-03	0,04444779	0,003192015	0,03528577
Iloraz szans z.jedn.	0,0003628807	1,18226		0	4,032443E+00	0,5071281	0	1,060837
-95%CL	9,57197400E-7	1,026096	1,370996E+09	0	1,621189E+00	0,2591248	0	1,003303
+95%CL	0,1375708	1,36219		0,00004660806	1,003004E+01	0,9924902	1,37033600E-34	1,12167
Iloraz szans zakr.		21,54648	6,474116E+02	0,0002323005	2,464948E+07	0,003029976	0,0000001149494	22,87554
-95%CL		1,603841	2,101539E+00	1,05692900E-7	3,641789E+02	9,79126900E-6	2,41586700E-12	1,190975
+95%CL		289,4618	1,994452E+05	0,5105688	1,668402E+12	0,9376467	0,005469406	439,3797

Tabela A 9 Wartości współczynników R Spearmana

	BMI	de Haan STI	Yeh DI	Organ BAND	van Geijn ID	Zugaib LTV	wiek
BMI	1	-0,094	-0,035	0,002	-0,085	-0,026	0,021
de Haan STI	-0,094	1,000	0,764	0,463	0,798	0,692	-0,304
Yeh DI	-0,035	*0,764	1,000	0,688	0,704	0,686	-0,242
Organ BAND	0,002	0,463	0,688	1,000	0,621	0,886	-0,376
van Geijn ID	-0,085	*0,798	*0,704	0,621	1,000	0,720	-0,381
Zugaib LTV	-0,026	0,692	0,686	*0,886	*0,720	1,000	-0,387
wiek	0,021	-0,304	-0,242	-0,376	-0,381	-0,387	1,000

*- oznaczono zmienne silnie skorelowane ($R > 0,7$), kolor czerwony oznacza wystąpienie istotnej statystycznie korelacji

3. pacjenci po przebytych zawale mięśnia sercowego, chorujący na nadciśnienie tętnicze i cukrzyce

zmienne BMI, de Haana STI

Model: Regr. logistyczna (logit) N zer: 62 jedynek: 18 (OBLICZENIA KONCOWE po 2 korekcje styczen201 Zmn. zal.: zawał, cukrzyca i nadciśnienie/zdrowy Strata: Największe prawd. bł.średnkw.skal. Całkowita strata: 31,417473069 Chi2(2)=22,471 p=,00001			
N=80	Stała B0	BMI	pr1_de Hann STI
Ocena	-9,609928	0,3493929	-86,11732
Błąd standard.	2,918515	0,1056159	41,4727
t(77)	-3,292746	3,308146	-2,076482
poziom p	0,001499877	0,001429566	0,04118587
-95%CL	-15,42143	0,1390847	-168,7
+95%CL	-3,798422	0,5597011	-3,534628
Chi-kwadrat Walda	10,84218	10,94383	4,311779
poziom p	0,0009932163	0,000940194	0,03785733
Iloraz szans z.jedn.	0,00006705966	1,418206	3,9785080E-38
-95%CL	2,00704300E-7	1,149221	0
+95%CL	0,02240609	1,750149	0,02916961
Iloraz szans zakr.		373,6804	0,01761679
-95%CL		10,56937	0,0003663097
+95%CL		13211,48	0,8472376

Tabela A 10 Wartość współczynnika R Spearmana

Para zmiennych	Korelacja porządku rang Spearmana Oznaczone wsp. korelacji są istotne z p <0,05			
	N ważnych	R Spearman	t(N-2)	p
BMI & de Haan STI	80	-0,231	-2,10029	0,039

kolor czerwony oznacza wystąpienie istotnej statystycznie korelacji

4. pacjenci chorujący na nadciśnienie tętnicze i cukrzycę

zmienne BMI, de Hanna STI, de Hanna LTI, Yeh DI, Yeh II, van Geijn ID, Zugaib LTV

N=104	Model: Regr. logistyczna (logit) N zer: 62 jedynek: 42 (OBLICZENIA KONCOWE po 2 korekcje styczen2007) Zmn. zal.: nadciśnienie i cukrzyca/zdrowy Strata: Największe prawd. bł.średnkw.skal. Całkowita strata: 46,682048947 Chi2(7)=46,940 p=,00000							
	Stała B0	BMI	pr1_de Hann STI	pr1_de Hann LTI	pr1_Yeh DI	pr1_Yeh II	pr1_van Geijn ID	pr1_Zugaib LTV
Ocena	-8,860357	0,3548655	87,11671	0,04655077	-350,959	297,6677	0,6238912	-504,5872
Błąd standard.	2,454189	0,08787324	91,81692	0,03727941	195,7132	143,6759	0,415694	234,0814
t(96)	-3,610299	4,03838	0,9488088	1,248699	-1,793232	2,0718	1,500842	-2,155606
poziom p	0,0004884305	0,0001083509	0,3451003	0,2148116	0,07608487	0,0409658	0,1366782	0,03361449
-95%CL	-13,73188	0,1804385	-95,13844	-0,02744827	-739,4467	12,47331	-0,2012549	-969,2352
+95%CL	-3,98883	0,5292925	269,3719	0,1205498	37,52853	582,8622	1,449037	-39,93936
Chi-kwadrat Walda	13,03426	16,30851	0,9002383	1,55925	3,21568	4,292355	2,252528	4,646637
poziom p	0,0003063109	0,0000539471	0,342725	0,2117842	0,07294532	0,03829209	0,1334062	0,03112187
Iloraz szans z.jedn.	0,0001419044	1,425989		1,047651	0		1,866176	0
-95%CL	1,08722300E-6	1,197742	0	0,972925	0	261270,1	0,817704	0
+95%CL	0,01852136	1,697731		1,128117			4,259013	4,5139490E-18
Iloraz szans zakr.		529,2712	57,45069	1297,03	3,26236600E-11		242,7851	7,2021780E-16
-95%CL		24,26101	0,01198694	0,01460412	8,04883700E-23	6,209576	0,1700493	8,1940340E-30
+95%CL		11546,43	275348,2	115192500	13,22307		346632,7	0,06330382

Tabela A 11 Wartości współczynników R Spearmana

	BMI	de Haan STI	de Haan LTI	Yeh DI	Yeh II	van Geijn ID	Zugaib LTV
BMI	1,000	-0,242	-0,249	-0,267	-0,226	-0,169	-0,217
de Haan STI	-0,242	1,000	0,759	0,889	0,730	0,815	0,780
de Haan LTI	-0,249	*0,759	1,000	0,742	0,871	0,658	0,928
Yeh DI	-0,267	*0,889	*0,742	1,000	0,866	0,819	0,840
Yeh II	-0,226	*0,730	*0,871	*0,866	1,000	0,743	0,975
van Geijn ID	-0,169	*0,815	0,658	*0,819	*0,743	1,000	0,783
Zugaib LTV	-0,217	*0,780	*0,928	*0,840	*0,975	*0,783	1,000

*- oznaczono zmienne silnie skorelowane ($R > 0,7$), kolor czerwony oznacza wystąpienie istotnej statystycznie korelacji

zmienne BMI, de Haana STI

Model: Regr. logistyczna (logit) N zer: 62 jedynek: 42 (OBLICZENIA KONCOWE po 2 korekcie styczen2011)			
Zmn. zal.: nadciśnienie i cukrzyca/zdrowy Strata: Największe prawd. bł.średnkw.skal.			
Całkowita strata: 51,783422976 Chi2(2)=36,738 p=,00000			
N=104	Stała B0	BMI	pr1_de Hann STI
Ocena	-8,408871	0,3274071	-68,86855
Błąd standard.	2,17145	0,07784337	29,141
t(101)	-3,872468	4,205973	-2,363287
poziom p	0,0001913972	0,0000563475	0,02002806
-95%CL	-12,71644	0,1729868	-126,6765
+95%CL	-4,101297	0,4818274	-11,06065
Chi-kwadrat Walda	14,996	17,69021	5,585128
poziom p	0,0001079532	0,0000260668	0,01811955
Iloraz szans z.jedn.	0,0002228813	1,387366	1,2324460E-30
-95%CL	3,00136200E-6	1,18885	0
+95%CL	0,01655119	1,61903	0,00001571878
Iloraz szans zakr.		325,7833	0,04066499
-95%CL		21,26746	0,002765718
+95%CL		4990,476	0,5979068

Tabela A 12 Wartość współczynnika R Spearmana

Para zmiennych	Korelacja porządku rang Spearmana Oznaczone wsp. korelacji są istotne z p <0,05			
	N ważnych	R Spearman	t(N-2)	p
BMI & de Haan STI	104	-0,242	-2,51522	0,013

kolor czerwony oznacza wystąpienie istotnej statystycznie korelacji

zmienne BMI, Zugaib LTV

Model: Regr. logistyczna (logit) N zer: 62 jedynek: 42 (OBLICZENIA KONCOWE po 2 korekcie styczn201 Zmn. zal.: nadciśnienie i cukrzyca/zdrowy Strata: Największe prawd. bł.średnkw.skala. Całkowita strata: 50,758225540 Chi2(2)=38,788 p=,00000			
	Stała B0	BMI	pr1_Zugaib LTV
N=104			
Ocena	-7,532538	0,3121077	-63,96084
Błąd standard.	2,149123	0,07580496	24,66875
t(101)	-3,504936	4,117246	-2,592788
poziom p	0,0006824912	0,0000784926	0,01093214
-95%CL	-11,79582	0,161731	-112,897
+95%CL	-3,269256	0,4624844	-15,02468
Chi-kwadrat Walda	12,28458	16,95172	6,722551
poziom p	0,0004573415	0,0000384387	0,009524466
Iloraz szans z.jedn.	0,0005353779	1,366302	1,66786500E-28
-95%CL	7,53599700E-6	1,175544	0
+95%CL	0,03803471	1,588014	2,98445500E-7
Iloraz szans zakr.		248,6005	0,01203792
-95%CL		17,4311	0,0004092509
+95%CL		3545,514	0,3540896

Tabela A 13 Wartości współczynników R Spearmana

Para zmiennych	Korelacja porządku rang Spearmana Oznaczone wsp. korelacji są istotne z p <0,05			
	N ważnych	R Spearman	t(N-2)	p
BMI & Zugaib LTV	104	-0,217	-2,24227	0,027

kolor czerwony oznacza wystąpienie istotnej statystycznie korelacji

zmienne Dalton MABB

Model: Regr. logistyczna (logit) N zer: 62 jedynek: 42 (OBLICZENIA KONCOWE po 2 korekcie styczn201		
Zmn. zal.: nadciśnienie i cukrzyca/zdrowy Strata: Największe prawd. bł.średnkw.skala.		
Całkowita strata: 62,117177972 Chi2(1)=16,070 p=,00006		
N=104	Stała B0	pr1_Dalton MABB
Ocena	1,012675	-0,1703714
Błąd standard.	0,4347316	0,05013243
t(102)	2,329425	-3,398427
poziom p	0,0218076	0,0009679222
-95%CL	0,1503865	-0,2698089
+95%CL	1,874963	-0,07093399
Chi-kwadrat Walda	5,426219	11,54931
poziom p	0,0198428	0,0006785706
Iloraz szans z.jedn.	2,752954	0,8433515
-95%CL	1,162283	0,7635254
+95%CL	6,520576	0,9315234
Iloraz szans zakr.		0,002046959
-95%CL		0,0000551741
+95%CL		0,07594214

Załącznik 3 Struktura drzew klasyfikacyjnych

1. Pacjenci po przebytych zawale mięśnia sercowego

Tabela A 14 Struktura drzewa klasyfikacyjnego dla pacjentów po przebytych zawale mięśnia sercowego

węzeł	gałąź lewostr.	gałąź prawostr.	n klas 0	n klas 1	klasa przewid.	stała podziału	zmienna podziału
1	2	3	62	84	1	27,1312	BMI
2	4	5	41	40	0	44,3567	de Haan LTI
3	6	7	21	44	1	29,1095	BMI
4	8	9	17	29	1	28,4384	de Haan LTI
5	10	11	24	11	0	112,2116	de Haan LTI
6	12	13	16	18	1	74,7413	de Haan LTI
7	14	15	5	26	1	32,0183	BMI
8	16	17	7	18	1	20,6594	de Haan LTI
9	18	19	10	11	1	23,0554	BMI
10	20	21	23	11	0	98,3771	de Haan LTI
11			1	0	0		
12	22	23	12	14	1	48,4495	de Haan LTI
13			4	4	0		
14	24	25	4	11	1	29,4431	BMI
15			1	15	1		
16			2	9	1		
17	26	27	5	9	1	26,0503	BMI
18			2	4	1		
19			8	7	0		
20			23	9	0		
21			0	2	1		
22	28	29	7	12	1	27,9705	BMI
23			5	2	0		
24			1	0	0		
25			3	11	1		
26			3	8	1		
27			2	1	0		
28	30	31	4	7	1	27,3721	BMI
29	32	33	3	5	1	28,4507	BMI

węzeł	gałąź lewostr.	gałąź prawostr.	n klas 0	n klas 1	klasa przewid.	stała podziału	zmienna podziału
30			2	1	0		
31			2	6	1		
32			2	1	0		
33			1	4	1		

2. Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie

Tabela A 15 Struktura drzewa klasyfikacyjnego dla pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie

węzeł	gałąź lewostr.	gałąź prawostr.	n klas 0	n klas 1	klasa przewid.	stała podziału	zmienna podziału
1	2	3	62	33	0	29,63616	BMI
2	4	5	59	22	0	0,02375	Zugaib LTV
3			3	11	1		
4	6	7	21	16	0	0,01485	Yeh DI
5	8	9	38	6	0	0,01105	Yeh DI
6	10	11	21	11	0	4,27545	Organ BAND
7			0	5	1		
8			3	4	1		
9	12	13	35	2	0	57,50000	wiek
10	14	15	21	9	0	53,50000	wiek
11			0	2	1		
12			27	0	0		
13	16	17	8	2	0	59,50000	wiek
14			7	0	0		
15	18	19	14	9	0	0,01565	de Haan STI
16			1	2	1		
17			7	0	0		
18	20	21	14	7	0	2,12300	van Geijn ID
19			0	2	1		
20			4	6	1		
21			10	1	0		

3. Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie tętnicze i cukrzycę

Tabela A 16 Struktura drzewa klasyfikacyjnego dla pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie i cukrzycę

węzeł	gałąź lewostr.	gałąź prawostr.	n klas 0	n klas 1	klasa przewid.	stała podziału	zmienna podziału
1	2	3	62	18	0	30,96984	BMI
2	4	5	60	10	0	0,01518	de Haan STI
3			2	8	1		
4	6	7	25	8	0	0,00422	de Haan STI
5	8	9	35	2	0	0,04126	de Haan STI
6			0	1	1		
7			25	7	0		
8			35	1	0		
9			0	1	1		

4. Pacjenci chorujący na nadciśnienie tętnicze i cukrzycę

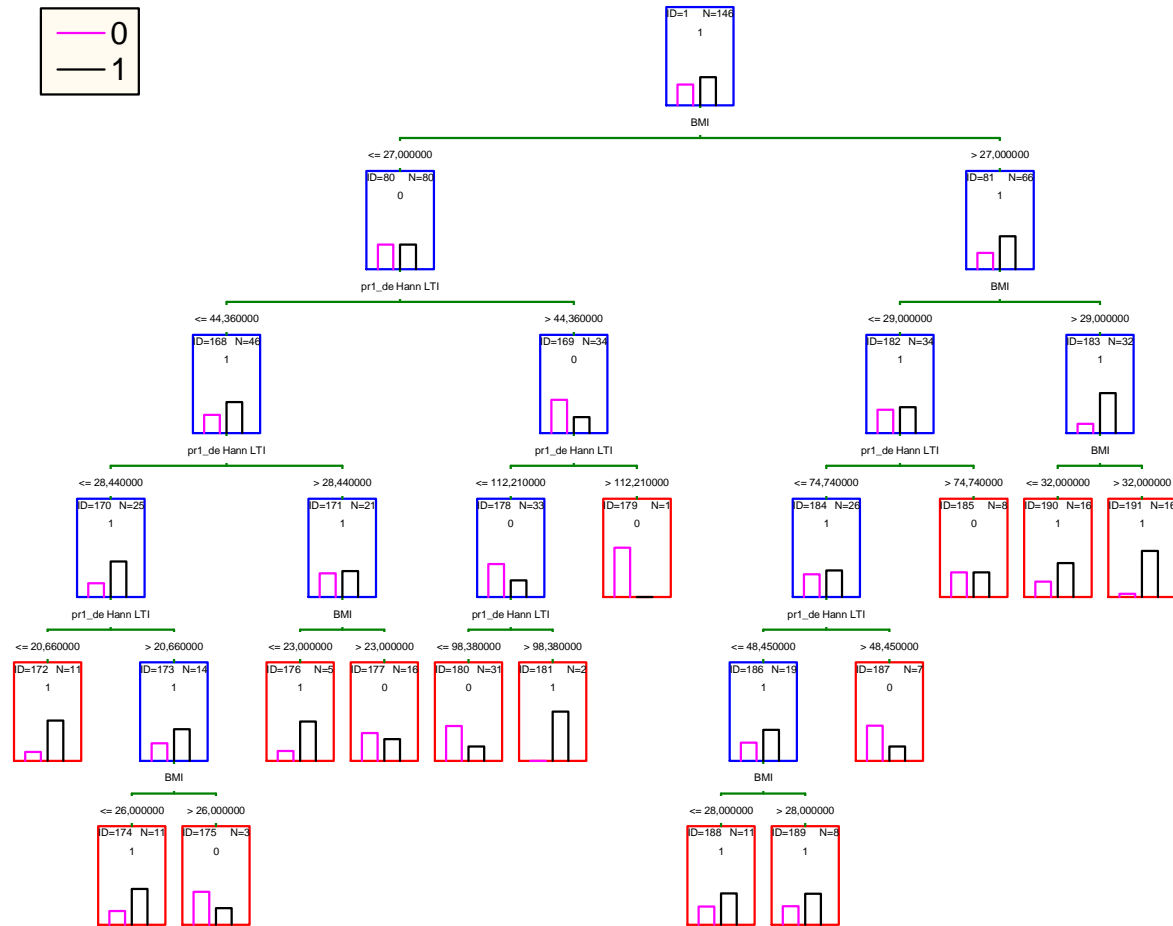
Tabela A 17 Struktura drzewa klasyfikacyjnego dla pacjentów chorujących na nadciśnienie i cukrzycę

gałąź	gałąź lewostr.	gałąź prawostr.	n klas 0	n klas 1	klasa przewid.	stała podziału	zmienna podziału
1	2	3	62	42	0	29,40211	BMI
2	4	5	58	20	0	0,02466	Zugaib LTV
3			4	22	1		
4	6	7	23	14	0	28,07173	BMI
5	8	9	35	6	0	0,03503	Zugaib LTV
6	10	11	20	10	0	25,35587	BMI
7			3	4	1		
8	12	13	18	4	0	0,02828	Zugaib LTV
9			17	2	0		
10			13	3	0		
11	14	15	7	7	0	26,84778	BMI

gałąź	gałąź lewostr.	gałąź prawostr.	n klas 0	n klas 1	klasa przewid.	stała podziału	zmienna podziału
12			12	2	0		
13	16	17	6	2	0	0,03116	Zugaib LTV
14			3	4	1		
15			4	3	0		
16			1	2	1		
17			5	0	0		

Załącznik nr 4 Struktura oraz macierze błędnych klasyfikacji dla drzew interakcyjnych

1. Pacjenci po przebytych zawale mięśnia sercowego



Wykres A 1 Struktura interakcyjnego drzewa klasyfikacyjnego – interpretacja graficzna (0 – nie, 1– tak)

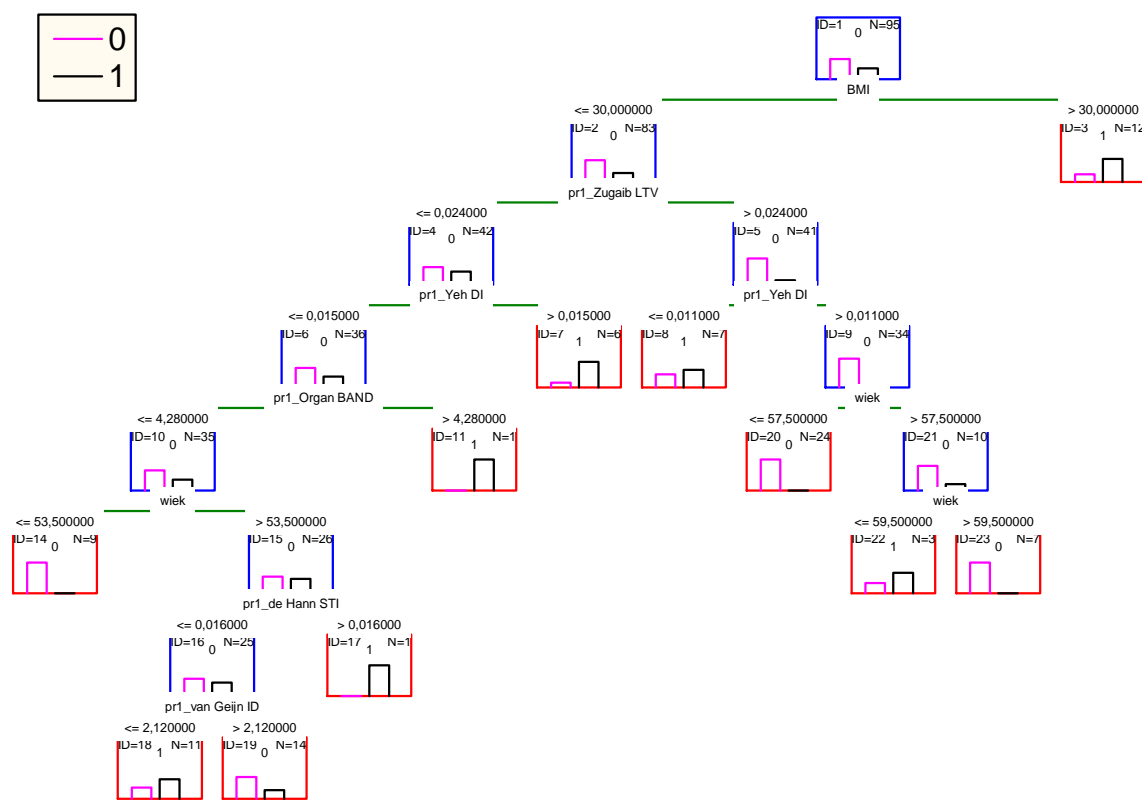
Tabela A 18 Macierz błędnych klasyfikacji pacjentów po przebytych zawale mięśnia sercowego i osób z grupy kontrolnej w modelu drzew interakcyjnych

Błędne klasyfikacje dla próby uczącej N=146		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	43	23
Przewidywane 1	19	61

Tabela A 19 Porównanie klasyfikacji pacjentów po przebytych zawale mięśnia sercowego i pacjentów z grupy kontrolnej dla drzew klasyfikacyjnych i interakcyjnych uzyskanej na zbiorze uczącym

	Drzewa klasyfikacyjne	Drzewa interakcyjne
% poprawnie zaklasyfikowanych przypadków	73,29	71,23
czułość	70,24	72,62
swoistość	77,42	69,35

2. Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie



Wykres A 2 Struktura interakcyjnego drzewa klasyfikacyjnego – interpretacja graficzna (0 – nie, 1– tak)

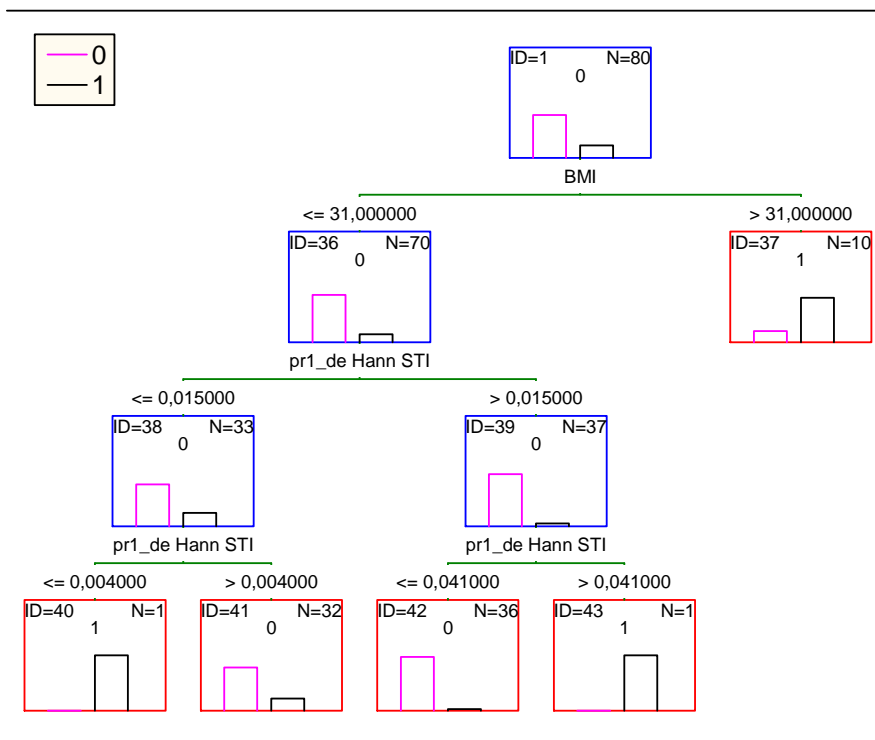
Tabela A 20 Macierz błędnych klasyfikacji pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie oraz osób z grupy kontrolnej w modelu drzew interakcyjnych

Błędne klasyfikacje dla próby uczącej N=95		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	50	4
Przewidywane 1	12	29

Tabela A 21 Porównanie klasyfikacji pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie tętnicze oraz pacjentów z grupy kontrolnej dla drzew klasyfikacyjnych i interakcyjnych uzyskanej na zbiorze uczącym

	Drzewa klasyfikacyjne	Drzewa interakcyjne
% poprawnie zaklasyfikowanych przypadków	87,37	83,16
czułość	96,97	87,88
swoistość	82,26	80,65

3. Pacjenci po przebytych zawale mięśnia sercowego i chorujący na nadciśnienie tętnicze i cukrzycę



Wykres A 3 Struktura interakcyjnego drzewa klasyfikacyjnego – interpretacja graficzna (0 – nie, 1– tak)

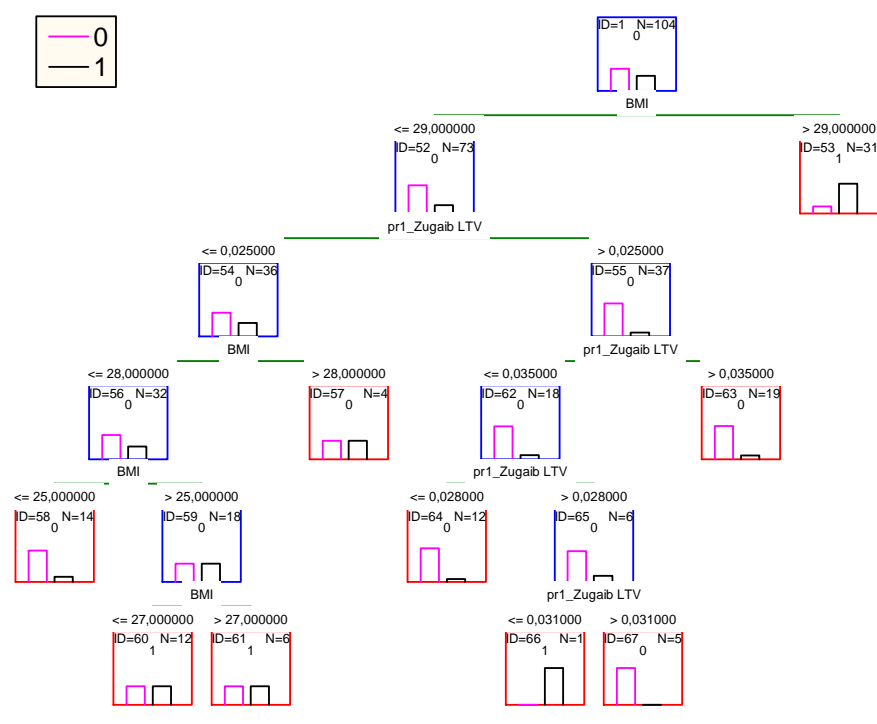
Tabela A 22 Macierz błędnych klasyfikacji pacjentów po przebytych zawale mięśnia sercowego i chorujących na nadciśnienie i cukrzycę oraz osób z grupy kontrolnej w modelu drzew interakcyjnych

Błędne klasyfikacje dla próby uczącej N=80		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	60	8
Przewidywane 1	2	10

Tabela A 23 Porównanie klasyfikacji pacjentów po przebytych zawałach mięśnia sercowego i chorujących na nadciśnienie tętnicze i cukrzycę oraz pacjentów z grupy kontrolnej dla drzew klasyfikacyjnych i interakcyjnych uzyskanej na zbiorze uczącym

	Drzewa klasyfikacyjne	Drzewa interakcyjne
% poprawnie zaklasyfikowanych przypadków	87,50	87,50
czułość	55,56	55,56
swoistość	96,77	96,77

4. Pacjenci chorujący na nadciśnienie tętnicze i cukrzycę



Wykres A 4 Struktura interakcyjnego drzewa klasyfikacyjnego – interpretacja graficzna (0 – nie, 1– tak)

Tabela A 24 Macierz błędnych klasyfikacji pacjentów chorujących na nadciśnienie i cukrzycę oraz osób z grupy kontrolnej w modelu drzew interakcyjnych

Błędne klasyfikacje dla próby uczącej N=104		
	Obserwowane 0	Obserwowane 1
Przewidywane 0	47	7
Przewidywane 1	15	35

Tabela A 25 Porównanie klasyfikacji pacjentów chorujących na nadciśnienie tętnicze i cukrzycę oraz pacjentów z grupy kontrolnej dla drzew klasyfikacyjnych i interakcyjnych uzyskanej na zbiorze uczącej

	Drzewa klasyfikacyjne	Drzewa interakcyjne
% poprawnie zaklasyfikowanych przypadków	79,81	78,85
czułość	76,19	83,33
swoistość	82,26	75,81