

Oświadczam, że niniejsza praca, stanowiąca podstawę do uznania osiągnięcia efektów uczenia się z przedmiotu Sieci neuronowe w zastosowaniach biomedycznych, została wykonana przeze mnie samodzielnie.

Ewa Mergo 300054

Karolina Olszewska 299697

Politechnika Warszawska

Wydział Elektroniki i Technik Informacyjnych

Sieci neuronowe w zastosowaniach biomedycznych - Projekt 1

Temat 25: Przewidywanie przewlekłej niewydolności nerek za pomocą sieci MLP

Wykonawca: Ewa Mergo, Karolina Olszewska - zespół 24

Opiekun projektu: dr inż. Bogumił Konarzewski

Warszawa 14.04.2021 r

1. Opis badanej choroby

Przewlekła niewydolność nerek jest chorobą obejmującą każde długo utrzymujące się uszkodzenie nerek. Nieprawidłowości w funkcjonowaniu nerek można wykryć na podstawie badań moczu, w badaniach obrazowych nerek czy poprzez zauważenie zmniejszonego przesączu kłębuszkowego.

W zależności od przyczyny przewlekłej niewydolności nerek, choroba ta może mieć różne objawy. Jeśli niewydolność rozwinęła się w następstwie kłębuszkowego bądź śródmiąższowego zapalenia nerek, objawiać będzie się krwimoczem, białkomoczem, wysokim ciśnieniem oraz obrzękami. Natomiast jeśli przyczyną jest cukrzyca lub nadciśnienie tętnicze, w moczu wykryć można albuminę, objawami będzie także częstsze oddawanie moczu i podwyższone ciśnienie. Innymi objawami mogą być także: utrata apetytu, trudności z oddychaniem, niedokrwistość, zaburzenia akcji serca, zapalenia osierdzia.

2. Wizualizacja i analiza danych.

Tabl. Dane

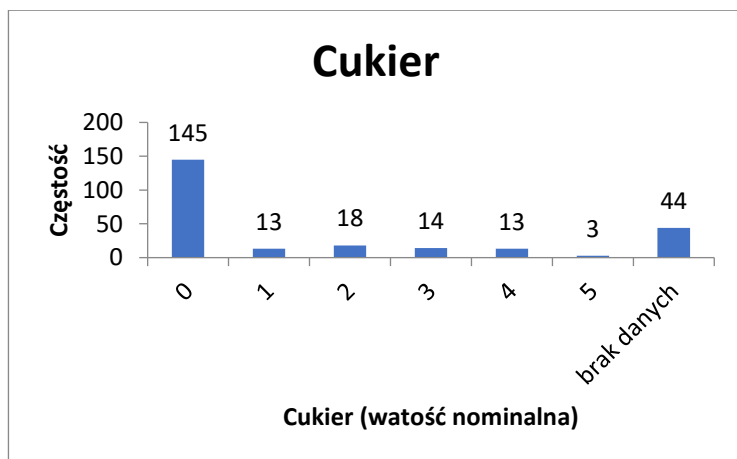
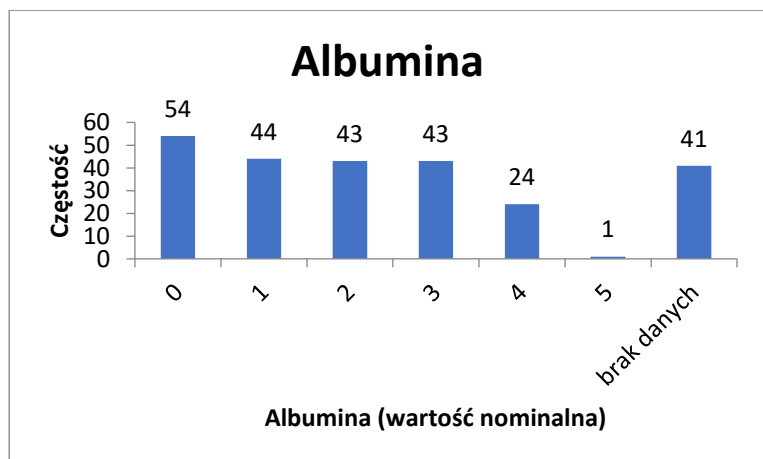
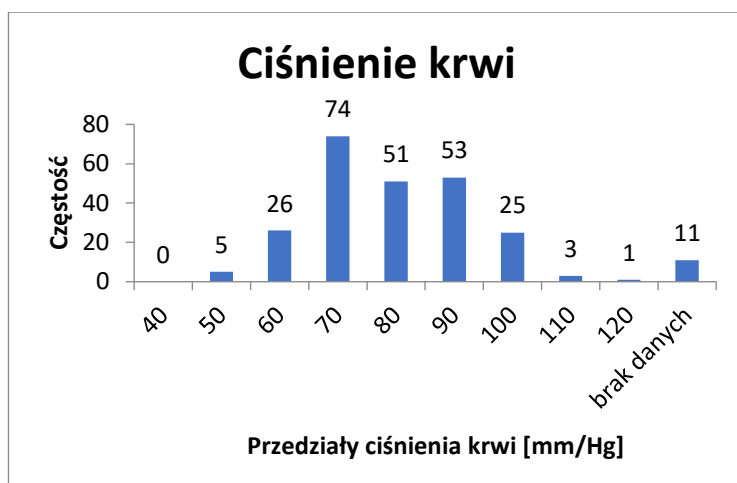
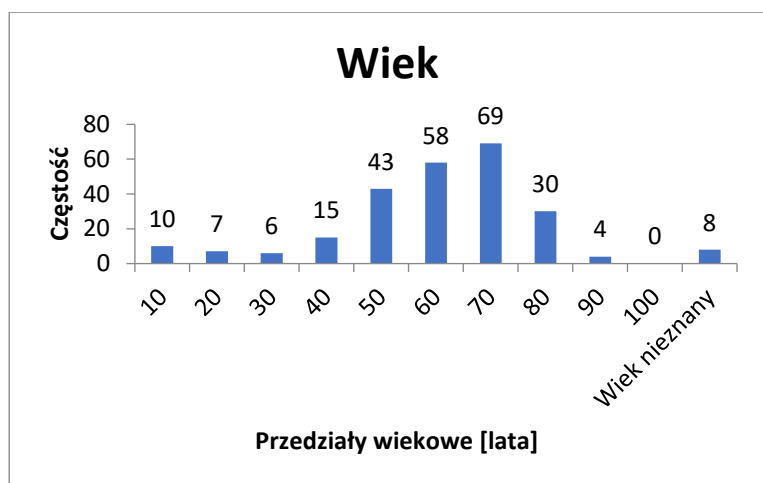
Nazwa	Typ	Jednostki numeryczne/ reprezentacja nominalna	Liczba niekompletnych danych
Wiek	Numeryczna	lata	9
Ciepłota ciała	Numeryczna	mm/Hg	12
Gęstość	Nominalna	1.005,1.010,1.015,1.020,1.025	47
Albumina	Nominalna	0,1,2,3,4,5	46
Cukier	Nominalna	0,1,2,3,4,5	49
Czerwone krwinki	Nominalna	w normie, poza normą	152
Komórki ropne	Nominalna	w normie, poza normą	65
Grudki komórek ropnych	Nominalna	występują, nie występują	4
Bakterie	Nominalna	występują, nie występują	4
Stężenie glukozy we krwi	Numeryczna	mgs/dl	44
Mocznik we krwi	Numeryczna	mgs/dl	19
Kreatynina w surowicy	Numeryczna	mgs/dl	17
Sód	Numeryczna	mEq/L	87
Potas	Numeryczna	mEq/L	88
Hemoglobina	Numeryczna	gms	52
Objętość komórek wypełnionych	Numeryczna	-	70
Liczba białych krwinek	Numeryczna	cells/cumm	105
Liczba czerwonych krwinek	Numeryczna	millions/cmm	130
Nadciśnienie tętnicze	Nominalna	tak, nie	2
Cukrzyca	Nominalna	tak, nie	2
Choroba wieńcowa	Nominalna	tak, nie	2
Apetyt	Nominalna	dobry, słaby	1
Obrzęk stóp	Nominalna	tak, nie	1
Anemia	Nominalna	tak, nie	1
Klasa	Nominalna	ckd, notckd	0

Liczba niekompletnych danych w przypadku cech : liczba czerwonych krwinek, liczba białych krwinek, czerwone krwinki przekracza 25% liczby wszystkich przypadków dlatego tych cech nie będziemy brać pod uwagę. (cechy skreślone w tabeli).

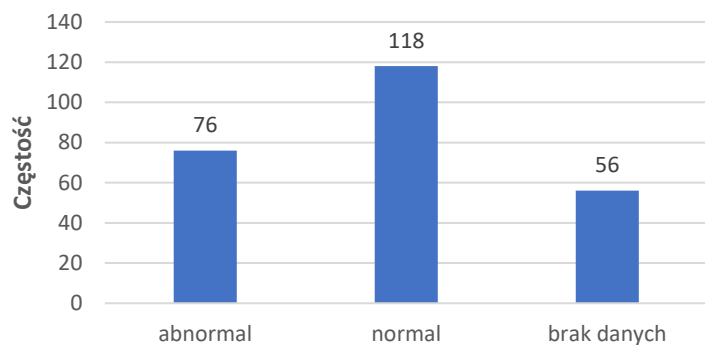
Tab2. Wybrane cechy do dalszych badań.

Nazwa	Zdrowy		Chory	
	Wartość średnia	Odchylenie standardowe	Wartość średnia	Odchylenie standardowe
Wiek	46,5	15,6	54,5	17,4
Ciepłota krwi	71,4	8,5	79,6	15,2
Albumina	-	-	-	-
Cukier	-	-	-	-
Komórki ropne	-	-	-	-
Bakterie	-	-	-	-
Stężenie glukozy we krwi	107,7	18,5	175,4	91,9
Mocznik we krwi	32,8	11,4	72,4	58,5
Kreatynina w surowicy	0,9	0,3	4,4	6,9
Hemoglobina	15,2	1,3	10,6	2,2
Nadciśnienie tętnicze	-	-	-	-
Cukrzyca	-	-	-	-
Choroba wieńcowa	-	-	-	-
Apetyt	-	-	-	-
Obrzęk stóp	-	-	-	-
Klasa	-	-	-	-

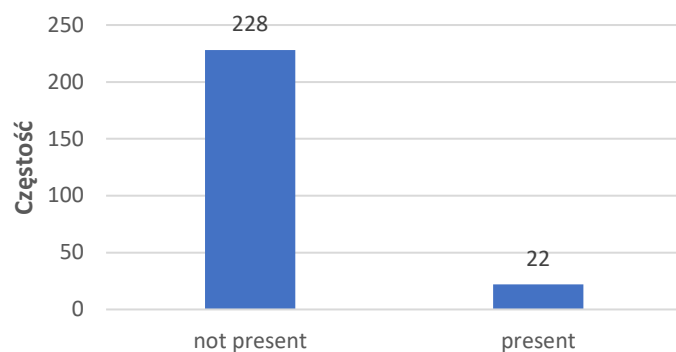
Analiza statystyczna przeprowadzona dla osób chorych na przewlekłe zapalenie nerek:



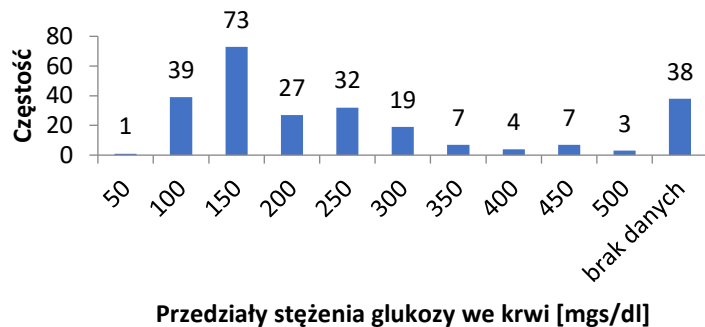
Komórki ropne



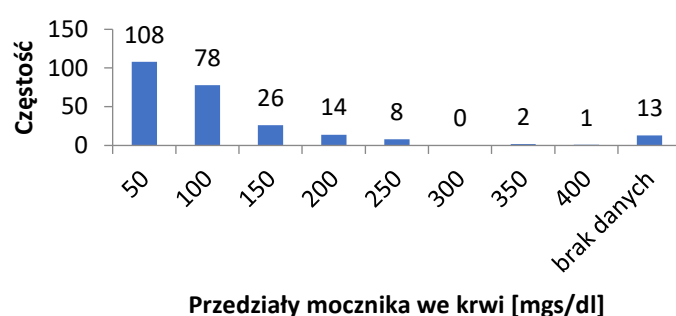
Bakterie



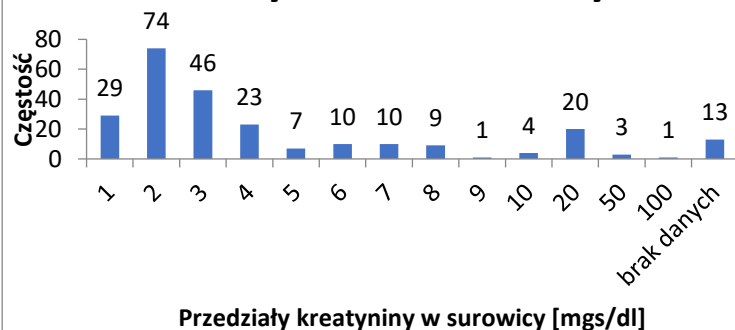
Glukoza



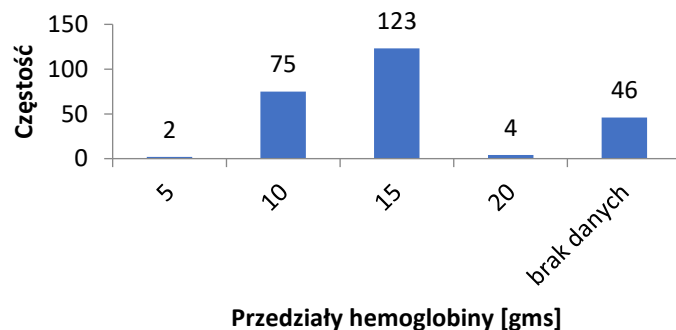
Mocznik



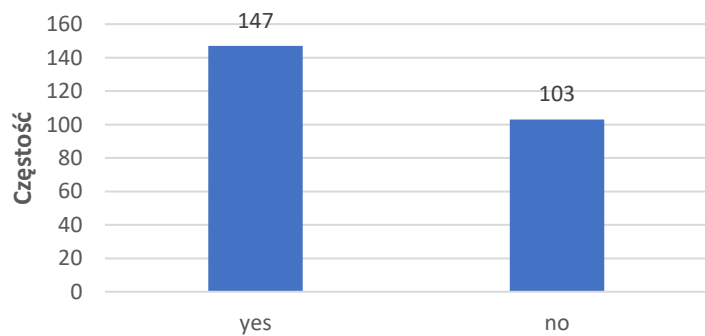
Kreatynina w surowicy



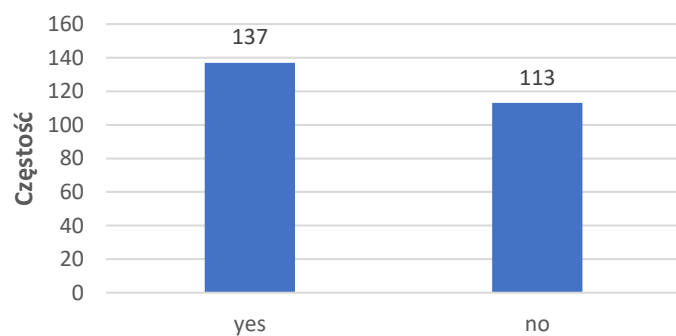
Hemoglobina



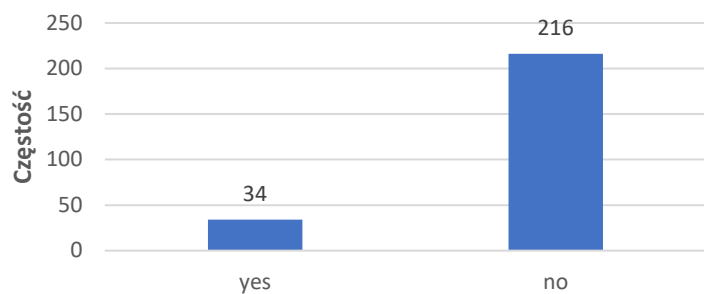
Nadciśnienie tętnicze



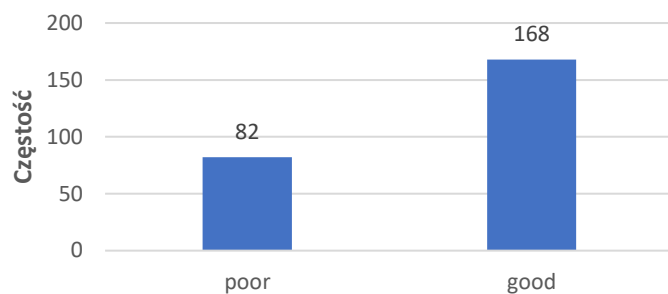
Cukrzyca



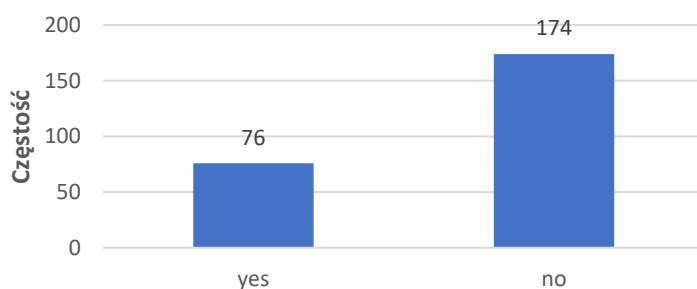
Choroba wieńcowa



Apetyt

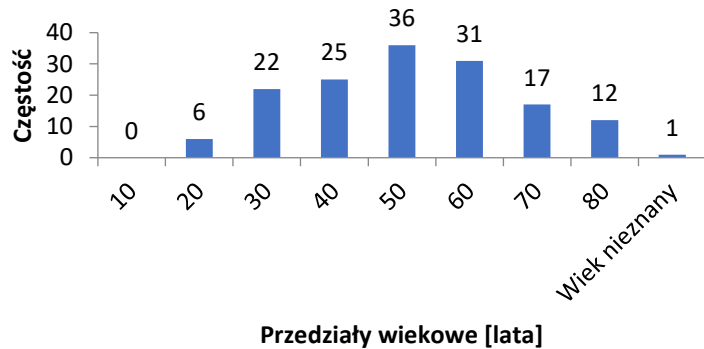


Obrzęk stóp

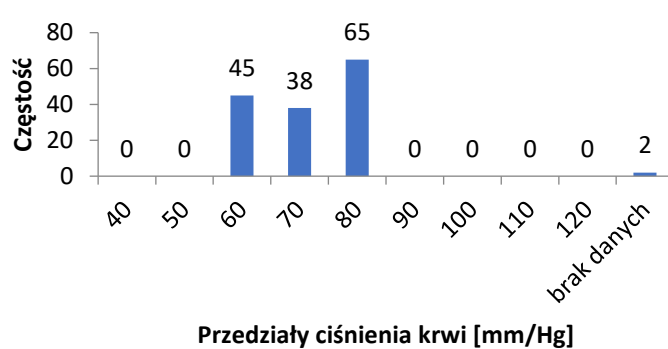


Punktem odniesienia są wyniki analizy statystycznej przeprowadzanej na osobach zdrowych:

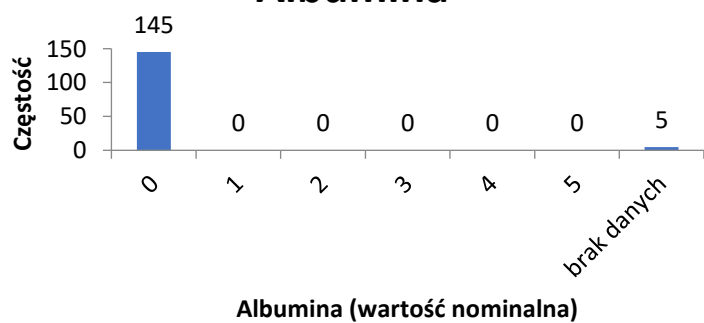
Wiek



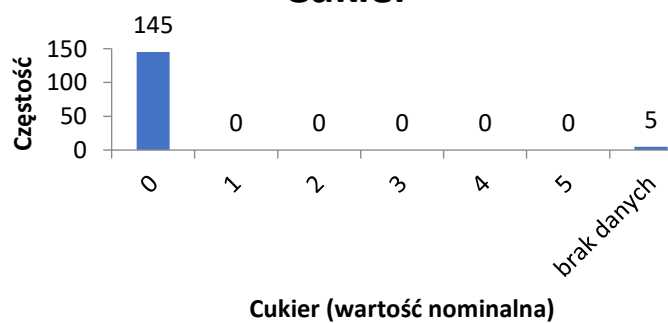
Ciśnienie krwi



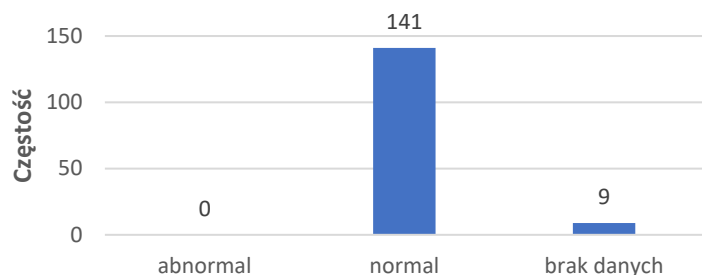
Albumina



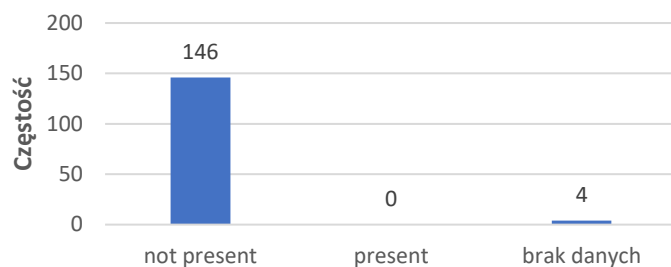
Cukier



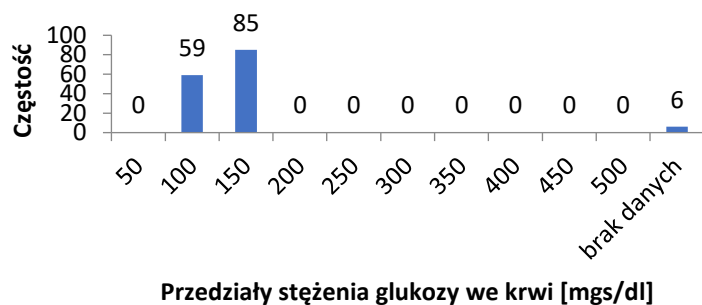
Komórki ropne



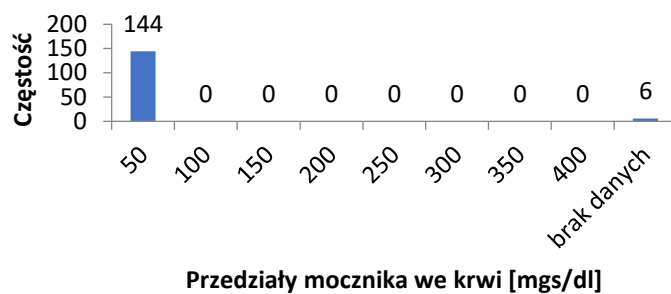
Bakterie



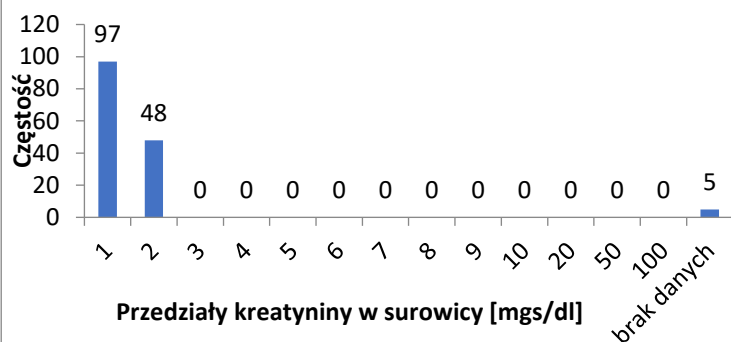
Glukoza



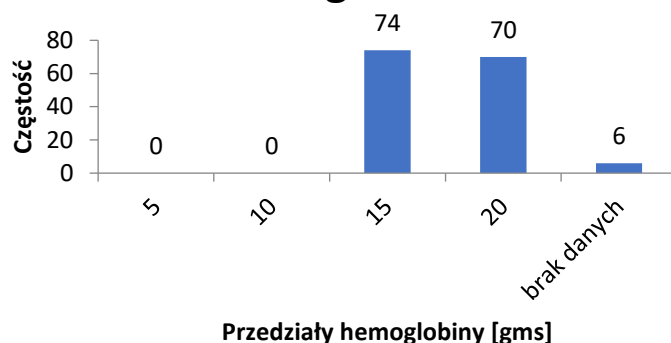
Mocznik



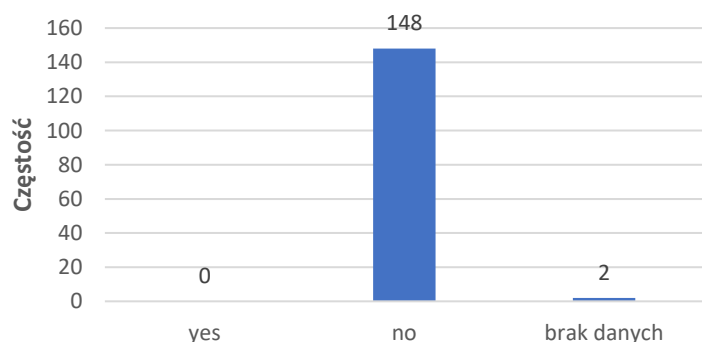
Kreatynina w surowicy



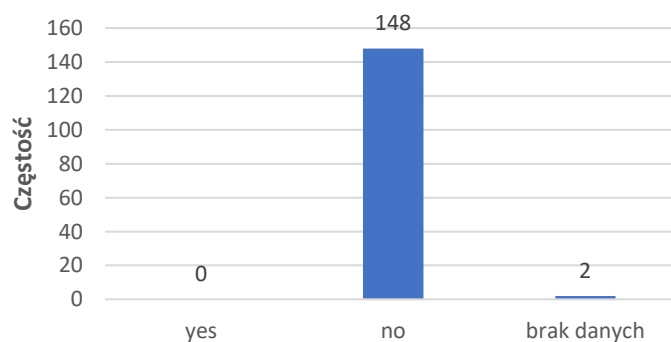
Hemoglobina

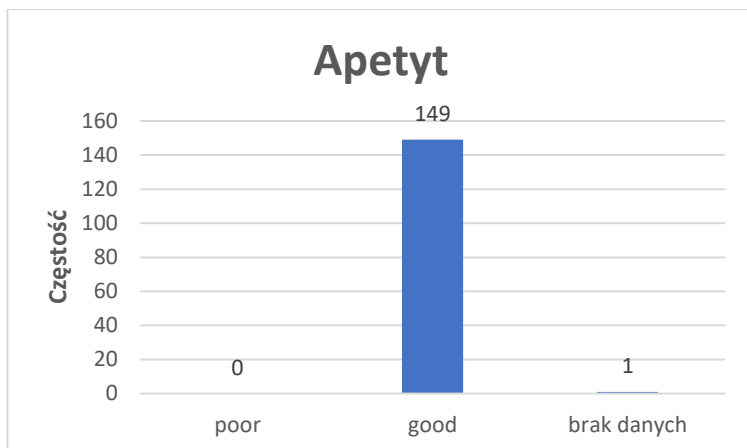
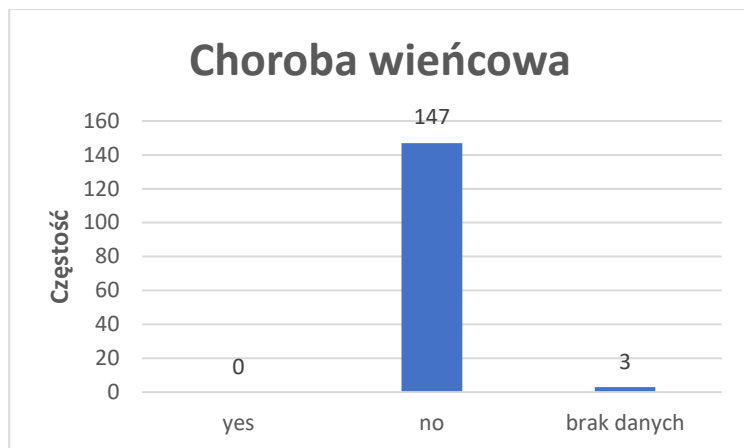


Nadciśnienie tętnicze



Cukrzyca





Analiza powyższych cech wskazuje, że wartości wybranych do analizy cech odbiegają od normy w przypadku przewlekłej niewydolności nerek.

3. Przetwarzanie danych przed podaniem ich na wejście sieci

- Proponowana metoda wstępnego przetwarzania danych i postępowanie w przypadku danych niekompletnych

Dane zostaną wstępnie przetworzone poprzez standaryzację:

$$z = \frac{x - \mu}{\sigma}$$

z – element po standaryzacji, x – element niestandaryzowany, μ – średnia arytmetyczna, σ – odchylenie standardowe

Natomiast dane niekompletne zostaną uzupełnione średnią wartością dla danej cechy.

- Kodowanie danych nienumerycznych

Nazwa	Reprezentacja nominalna	Zmiana	Reprezentacja nominalna	Zmiana
Komórki ropne	w normie	1	poza normą	0
Bakterie	nie występują		występują	
Nadciśnienie tętnicze	nie		tak	
Cukrzyca				
Choroba wieńcowa				
Obrzęk stóp				
Apetyt	dobry		słaby	
Klasa	notckd		ckd	

- Proponowany podział zbioru danych

Dane uczące stanowią 80% wszystkich danych, czyli 200 wektorów z klasy chorych i 120 z klasy zdrowych. Pozostałe 20% danych zostaną przeznaczone na dane testowe.

4. Koncepcja realizacji sieci neuronowej

Struktura sieci neuronowej:

- Warstwa wejściowa – 16 neuronów
- Warstwa ukryta – 4 neurony
- Warstwa wyjściowa – 1 neuron

Liczba neuronów w warstwie ukrytej została wyliczona z zależności:

$$N_{ukryte} = \sqrt{N_{WE} * N_{WY}}$$

N_{WE} – liczba neuronów warstwy wejściowej
 N_{WY} – liczba neuronów warstwy wyjściowej

Wybrana została liniowa funkcja jako funkcja aktywacji.

Zastosowany zostanie algorytm wstecznej propagacji błędów, który obejmuje następujące kroki:

- przyjęcie niewielkiej losowej wartości wag $w_{l,n}$ oraz $\omega_{m,l}$
- podanie na wejście losowo wybrany wektor uczący
- obliczenie pobudzenia neuronów warstwy ukrytej
- obliczenie stanu wyjść neuronów warstwy wyjściowej
- obliczenie pobudzenia neuronów warstwy wyjściowej
- obliczenie stanu wyjść neuronów warstwy wyjściowej
- obliczenie sygnału błędu dla warstwy wyjściowej
- obliczenie sygnału błędu dla warstwy ukrytej
- zmodyfikowanie wagi warstwy wyjściowej
- zmodyfikowanie wagi warstwy ukrytej
- sprawdzenie czy poziom błędu jest zadowalający – jeśli nie to należy wrócić do punktu 2