

Određivanje krvnog tlaka iz elektrokardiograma metodama strojnog učenja

Pap, Marko

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Electrical Engineering and Computing / Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:168:704257>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-03-14**



Repository / Repozitorij:

[FER Repository - University of Zagreb Faculty of Electrical Engineering and Computing repository](#)



SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 287

**ODREĐIVANJE KRVNOG TLAKA IZ
ELEKTROKARDIOGRAMA METODAMA STROJNOG
UČENJA**

Marko Pap

Zagreb, veljača 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 287

**ODREĐIVANJE KRVNOG TLAKA IZ
ELEKTROKARDIOGRAMA METODAMA STROJNOG
UČENJA**

Marko Pap

Zagreb, veljača 2024.

DIPLOMSKI ZADATAK br. 287

Pristupnik: **Marko Pap (1191239332)**

Studij: Računarstvo

Profil: Znanost o podacima

Mentor: prof. dr. sc. Darko Vasić

Zadatak: **Određivanje krvnog tlaka iz elektrokardiograma metodama strojnog učenja**

Opis zadatka:

Cilj ovog rada je razviti metodu strojnog učenja za procjenu krvnog tlaka iz elektrokardiograma. Pretražiti i proučiti javno dostupne baze elektrokardiograma i podataka o krvnom tlaku za veliki broj ispitanika. Razviti sučelje za pristup odabranoj bazi podataka, razviti algoritme digitalne predobradbe elektrokardiografskih signala (filtriranje, uklanjanje bazne linije i segmentiranje) te vremenski korelirati elektrokardiogram s neinvazivnim mjerenjem krvnog tlaka. Tako formiran skup podataka podijeliti u skupove za učenje, verifikaciju i validaciju. Odabrati prikladnu metodu strojnog učenja s osvrtom na stanje tehnologije u literaturi. Implementirati metodu predikcije krvnog tlaka u barem nekoliko kvalitativnih kategorija (npr. nizak, normalan, povišen i visok tlak) te usporediti upotrebu i točnost s komercijalno dostupnim neinvazivnim tlakomjerima.

Rok za predaju rada: 9. veljače 2024.

Dragi Mama i Tata prvo i najviše se želim zahvalit vama, za upornost od srednje škole pa sve do samog kraja. Nikad nije ni bilo upitno hoću li završiti ili ne ja sam to znao da budem, ali vas je nekada bilo malo teže za uvjeriti. Sada u miru možete znat da je to gotovo i da ste svoj posao obavili. Hvala vam !

Drugo bi se htio zahvalit svim prijateljima koji su ostavili trag u mojem putovanju i kroz neizmjernu potporu i duga učenja i zabavu pomogli bi da dosegнем ovaj uspjeh.

Također uzeo bih ovu priliku te se zahvalit svim Profesorima koji su mi predavali i prenijeli neprocjenjivo znanje ne samo teoretsko nego i životno koje će me pratit kroz cijeli život.

Zadnje, ali ne i najmanje, jedno veliko hvala Luka ti si mi oduvijek i zauvijek ćeš biti moj uzor, postavio si visoku ljestvicu ispred mene mislim da u ovom poglavlju života nisam bio daleko od nje i samo tako nastavi dizat ju i zajedno ćemo rast i bit uspješniji.

Hvala vam svima, Marko !

Sadržaj

Uvod	1
1. Radno okruženje	3
2. Baza izmjerenih signala EKG-a i tlaka	4
2.1. Analiza podataka	5
2.2. Prikaz i ocjena kvalitete signala	7
2.2.1. EKG signal ne zadovoljava osnovne kriterije.....	8
2.2.2. Signal krvnog tlaka izlazi iz okvira mjerenja	8
2.3. Finalni set podataka	8
3. Pred obrada signala	9
3.1. Filtriranje EKG signala i micanje osnovne linije signala	9
3.2. Normalizacija signala	9
4. Odabir značajki signala	11
5. Klase krvnog tlaka	15
6. Izrada glavne tablice	17
6.1. Podijele na skup za treniranje i testiranje.....	17
7. Izrada klasifikacijskog modela strojnog učenja	19
7.1. Stablo odlučivanja.....	19
7.2. Slučajna šuma	19
7.3. Potporni vektorski stroj (SVM)	20
7.4. Logistička regresija.....	20
7.5. Standardizacija numeričkih značajki	21
7.6. Balansiranje klasa	21
7.7. Optimizacija hiperparametara i treniranje modela	21
7.8. Evaluacija modela	22
8. Rezultati	23

8.1.	Klasifikacijski modeli 2 klase	23
8.2.	Klasifikacijski modeli 3 klase	25
8.3.	PCA analiza značajki	27
	Zaključak	28
	Literatura	29
	Sažetak.....	31
	Summary	32

Uvod

Krvni tlak [1] predstavlja silu kojom krv pritišće stijenke krvnih žila (arterija) dok cirkulira kroz tijelo. Krvni tlak izražava se obično dvjema vrijednostima: sistoličkim i dijastoličkim tlakom, mjerenim u milimetrima žive (mmHg). Sistolički tlak: To je viša od dvije vrijednosti i označava tlak u arterijama tijekom kontrakcije srčanog mišića (sistole), kada se krv pumpa u arterije. To je prvi broj u mjerenju krvnog tlaka. Dijastolički tlak: To je niža od dvije vrijednosti i predstavlja tlak u arterijama tijekom opuštanja srca (dijastole), kada se srce puni krvlju. To je drugi broj u mjerenju krvnog tlaka. Elektrokardiogram (EKG) [2] Elektrokardiografija je metoda registracije bioelektričkih potencijala nastalih električkom aktivnošću srca. Uređaji za registraciju napona srca nazivaju se elektrokardiografi, a zapisi tih napona na papiru, elektrokardiogrami.

U današnjem dobu, kada tehnološki napredak neprestano oblikuje područje medicinske dijagnostike, istraživači i zdravstveni stručnjaci tragaju za inovativnim pristupima koji bi unaprijedili preciznost i praktičnost mjerenja vitalnih parametara. Među tim parametrima, krvni tlak ima ključnu ulogu u procjeni kardiovaskularnog zdravlja pacijenata. Tradicionalne metode mjerenja krvnog tlaka često su vezane uz neinvazivne metode poput upotrebe auskultativne metode [3] na primjer manžete i stetoskopa, no u posljednje vrijeme sve veća pažnja usmjerena je prema integraciji naprednih tehnologija u dijagnostičke postupke. Ovaj rad posvećen je razvoju pristupa određivanju krvnog tlaka, koristeći elektrokardiografske signale i metode strojnog učenja. Elektrokardiografija (EKG) pruža uvide u električnu aktivnost srca, ali njezina primjena u procjeni krvnog tlaka predstavlja izazov koji zahtijeva integraciju digitalne predobrade signala, analize podataka i pravilnog koreliranja s neinvazivnim mjerenjem krvnog tlaka. Jedan od ključnih izazova u ovom području istraživanja jest pristup kvalitetnim podacima. Stoga, u sklopu ovog rada, istražiti ćemo javno dostupne baze elektrokardiograma i podataka o krvnom tlaku za veliki broj pacijenata. Analiza tih podataka omogućit će nam dublje razumijevanje varijacija elektrokardiografskih signala u kontekstu krvnog tlaka, a time i bolje oblikovanje algoritama za digitalnu predobradu i analizu. Digitalna predobrada elektrokardiografskih signala, poput filtriranja, uklanjanja bazne linije i segmentacije, ključna je faza u procesu analize. Ove tehnike omogućuju izdvajanje relevantnih informacija iz sirovih EKG signala, čime se stvara temelj za daljnju obradu podataka. Kroz razvoj algoritama za digitalnu predobradu, ovaj rad nastoji optimizirati prijenos korisnih informacija iz elektrokardiograma u kontekstu procjene krvnog tlaka. Nadalje, kako bismo validirali razvijenu metodu, podijelit ćemo formirani skup podataka u skupove za učenje i validaciju. Ovaj pristup omogućit će procjenu učinkovitosti metode na različitim podskupovima podataka, čime ćemo osigurati pouzdanost i generalizaciju rezultata. U odabiru metode strojnog učenja, posebna pažnja bit će posvećena stanju tehnologije u literaturi. Analizirat ćemo različite pristupe i modele koji su se pokazali uspješnima u sličnim zadacima te odabrati onaj koji najbolje odgovara specifičnostima našeg problema. Konačno, implementacija metode predikcije krvnog tlaka u različite kvalitativne kategorije (sniženi, normalni i povišeni) omogućit će usporedbu s komercijalno dostupnim neinvazivnim tlakomjerima. Cilj je ne samo razviti inovativan pristup, već i potvrditi njegovu upotrebljivost i točnost u usporedbi s već ustanovljenim standardima. Ovaj rad stoga predstavlja korak prema integraciji modernih tehnologija strojnog učenja u domenu

medicinske dijagnostike, pridonoseći unapređenju kvalitete skrbi za pacijente i omogućujući preciznije procjene krvnog tlaka temeljene na elektrokardiografskim signalima.

Za izradu ovog rada inspiracija je crpljena iz prethodnih istraživanja, posebice iz rada [4]. Rad uvodi pristup temeljen na strojnom učenju za procjenu krvnog tlaka koristeći isključivo EKG signale. Podaci od 51 pacijenta, što rezultira s 3129 EKG segmenata, analizirani su primjenom analize kompleksnosti kako bi se izdvojile karakteristike. Ove karakteristike uključivale su pokretljivost signala, kompleksnost, fraktalnu dimenziju, entropiju, autokorelaciju i dob. Razvijen je model koji kombinira klasifikaciju (Kombiniranje algoritama strojnog učenja) i regresijske modele za predviđanje krvnog tlaka. Model klasificira EKG signale u kategorije normalnog, prehipertenzije i hipertenzije te predviđa konkretne vrijednosti krvnog tlaka. Prosječna apsolutna pogreška u predviđanjima bila je 8.64 mmHg za sistolički, 18.20 mmHg za dijastolički i 13.52 mmHg za srednji arterijski tlak. Kalibracija korištenjem distribucije vjerojatnosti dodatno je poboljšala točnost, posebno za dijastolički tlak. Performanse metode bile su bliske certificiranim medicinskim uređajima, što ukazuje na njezin potencijal za neinvazivno kontinuirano praćenje krvnog tlaka. Izazovi uključuju varijabilnost individualnih fizioloških odgovora i potrebu za daljnjom validacijom. Sveukupno, studija sugerira da EKG signali mogu biti održiv izvor za procjenu krvnog tlaka upotrebom naprednih tehnika strojnog učenja. Također je pokazano da se može postići velika klasifikacijska točnost krvnog tlaka koristeći značajke izvučene iz signala fotopletizmografa (PPG) i elektrokardiograma (ECG) putem konvolucijske neuronske mreže (CNN) [5]. Dio pristupa analizi signala je potaknut inspiracijom iz literature [6].

1. Radno okruženje

Ovaj rad je u potpunosti implementiran kroz programski kod jezika Python verzije 3.11.3. u radnom okruženju Jupyter bilježnice.

Na Slici 1 vidimo popis svih biblioteka koje su se koristile u implementaciji ovog rada.

```
import pandas as pd
import numpy as np
import math
import matplotlib.pyplot as plt
import seaborn as sns
import scipy
from statistics import mean, median

import heartpy as hp
import wfdb as wf
from wfdb import processing

from scipy import signal
from scipy.interpolate import interp1d

from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import KFold, train_test_split
from sklearn.metrics import mean_squared_error
```

Slika 1. Python biblioteke

1. Pandas: Koristi se za obradu i analizu podataka. Ova biblioteka omogućava učinkovito manipuliranje i analizu velikih skupova podataka, te često služi za učitavanje, čišćenje, istraživanje i manipuliranje podacima.
2. NumPy: Osnovna biblioteka za numeričke proračune u Pythonu. Omogućava rad s velikim, višedimenzionalnim nizovima i matricama, uz širok spektar matematičkih funkcija za rad s tim nizovima.
3. Math: Standardna Python biblioteka za matematičke funkcije. Pruža pristup osnovnim matematičkim operacijama i funkcijama.
4. Matplotlib: Biblioteka za vizualizaciju podataka koja omogućava crtanje grafikona i dijagrama.
5. Seaborn: Zasnovan na Matplotlibu, Seaborn olakšava stvaranje lijepih, informativnih statističkih grafova.
6. SciPy: Biblioteka korištena za znanstvene i tehničke računske operacije.
7. Statistics: Standardna biblioteka za osnovne statističke operacije.
8. HeartPy: Specijalizirana biblioteka za analizu podataka otkucaja srca. Omogućava obradu signala otkucaja srca za izvlačenje korisnih metrika.
9. WFDB: Biblioteka koja se koristi za rad sa fiziološkim signalima, posebno onima iz PhysioNet baze podataka. Omogućava učitavanje, pisanje i obradu različitih fizioloških signala.
10. Sklearn: Ova biblioteka je ključna za strojno učenje u Pythonu. Pruža alate za modeliranje podataka, kao i za izgradnju modela.

2. Baza izmjerenih signala EKG-a i tlaka

Baza podataka je dostupna na web portalu PhysioNet.org (referenca), podaci se nalaze pod naslovom „*Autonomic Aging: A dataset to quantify changes of cardiovascular autonomic function during healthy aging*“ [7]. Javno je dostupna pod licencom „Open Data Commons Open Database License v1.0“ [8].

Podaci su dostavljeni u otvorenom WFDB standardnom formatu i imenovani su uzastopnim brojevima nakon nasumičnog poretka. Dodatne informacije o pacijentima pohranjene su u datoteci subject-info.csv. Osigurano je da se nijedan od pacijent ne može identificirati na temelju demografskih informacija, generalizirali su pojedinačnu dob u dobne skupine. Koristili su besplatni alat za anonimizaciju podataka ARX [9] s primijenjenim uvjetom anonimnosti $k=2$ i prosječnim rizikom ponovnog identificiranja ispod 5%.

Sve mjerenja su zabilježena na Odjelu za psihosomatsku medicinu i psihoterapiju Sveučilišne bolnice Jena. Studija je odobrena od strane Etičkog odbora Medicinskog fakulteta Sveučilišta Friedrich Schiller Jena. Sva istraživanja provedena su u skladu s relevantnim smjernicama i propisima. Informirani pismeni pristanak dobiven je od svih pacijenata.

Drugi Einthovenov odvod EKG-a sniman je frekvencijom uzorkovanja od 1000 Hz, ili pomoću MP150 sustava [10] koji omogućava snimanje i analizu fizioloških signala iz ljudi i životinja pomoću širokog spektra senzora i modula ili sustava Task Force Monitor [11] koji neinvazivno može kontinuirano mjeriti krvni tlak i hemodinamičke parametre. Prethodno gelirane elektrode od Ag/AgCl postavljene su prema Einthovenovom trokutu.

Kontinuirani krvni tlak sniman je neinvazivno pomoću tehnike vaskularnog rasterećenja. Ukratko, manžeta oko prsta kontrolira se kako bi održala konstantan tlak, dok se količina krvi bilježi putem fotopletizmografije [12]. U slučaju nepromjenjivog tlaka manžete tlakom manžete, dobivena količina krvi proporcionalna je krvnom tlaku u arterijama prsta. Snimljeni signal mapira se na brahijalni krvni tlak koji se mjeri oscilometrijski jednom tijekom inicijalizacije sustava. Sustav MP150 digitalizira signal koji se dobiva putem zasebnog monitora CNAP 500.

Mjerenja su obavljena u preglednoj sobi koja je bila kontrolirana na temperaturi od 22°C. Tijekom snimanja bilo je apsolutno tiho i potpuno zasjenjeno. Razina osvjetljenja održavana je konstantnom putem neizravnog izvora svjetlosti. Snimanje je započelo razgovorom s pacijentom. Zatim su mu objašnjena svrha i dizajn studije. Svaki od pacijenata u ovoj kolekciji dao je pismeni pristanak za sudjelovanje.

Nakon što su se sudionici udobno smjestili na stol za pregled, postavljeni su elektrodi i manžete. Za snimanje u mirovanju, sudionici su bili upućeni da izbjegavaju pokrete, zijevanje ili kašljanje.

Instruktor je pričekao nekoliko minuta da se pacijent smiri, provjerio kvalitetu dobivenih signala. U slučaju nedovoljne kvalitete signala, elektrodi i manžete su ponovno raspoređeni. Inače, snimanje je započelo. Prosječno trajanje snimanja bilo je 19 minuta (8 - 45 minuta) i nadzirao ga je instruktor.

Baza se sastoji od 1121 zapisa za svakog pacijenata jedan zapis koji se sastojao od: signala EKG-a, signala krvnog tlaka, dobne skupine, indeks tjelesne mase, spola i mjernih uređaja. Gdje su dobne skupine definirane na sljedeći način: 1 (18-19 godina), 2 (20-24 godine), 3

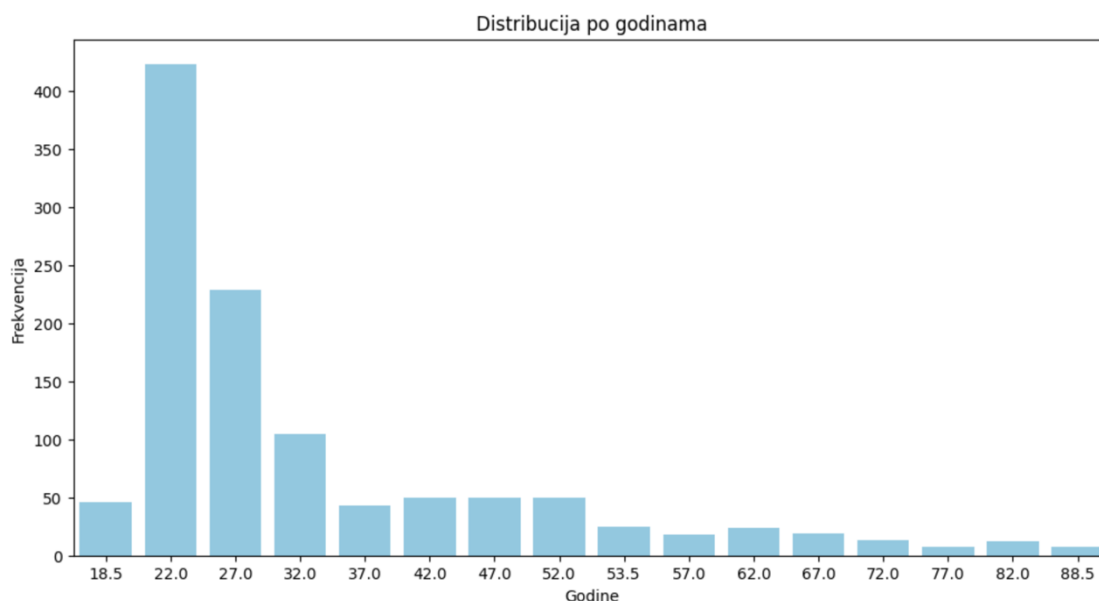
(25-29 godina), 4 (30-34 godine), 5 (35-39 godina), 6 (40 -44 godine), 7 (45-49 godina), 8 (50-54 godine), 9 (55-59 godina), 10 (60-64 godine), 11 (65-69 godina), 12 (70-74 godine), 13 (75-79 godina), 14 (80-84 godine), 15 (85-92 godine). Spol je kodiran 0 (muški) ili 1 (ženski). U Tablici 1 vidimo strukturu baze signala.

Tablica 1. Struktura baze signala

Kolona	Vrsta zapisa
EKG signal	Lista – float vrijednosti
Signal krvnog tlaka	Lista – float vrijednosti
Godine	Float
Spol	0 ili 1
BMI	Cijeli broj

2.1. Analiza podataka

Prije početka pred procesiranja signala potrebno je analizirati podatke te provjeriti dali nešto treba izbaciti ili dodati. Prvi korak koji je napravljen je to da su se dobne skupine kodirale u srednju vrijednost intervala koji zastupaju: 1 – 18.5 godina, 2 - 22 godine, 3 - 27 godina, 4 - 32 godine, 5 - 37 godina, 6 - 42 godine, 7 - 47 godina, 8 - 52 godine, 9 - 57 godina, 10 - 62 godine, 11 - 67 godina, 12 - 72 godine, 13 - 77 godina, 14 - 82 godine, 15 – 88.5 godine. Postoji 25 zapisa s nepostojećom vrijednosti, tim zapisima je dodijeljena srednja vrijednost ukupnog intervala 18.5-88.5 godina što je 53.5 godina. Razdioba je prikazana u obliku stupčastog grafa na Slici 1:



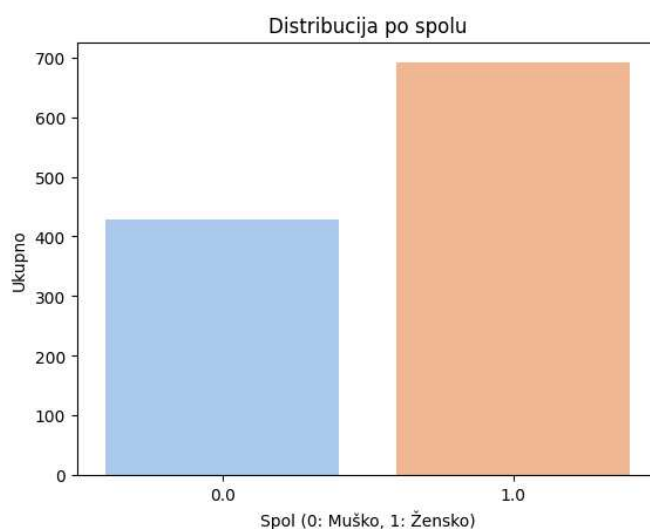
Slika 2. Distribucija pacijenata po godinama

Razdioba po spolu dana je u Tablici 2

Tablica 2. Frekvencija po spolu

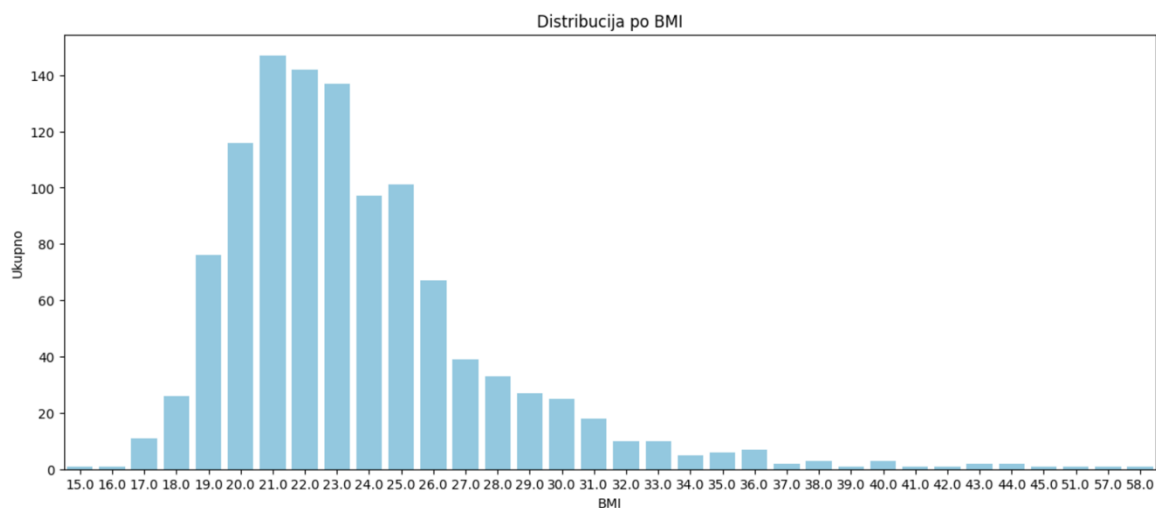
Spol	Frekvencija
0 - Muško	430
1 - Žensko	691

Distribucija po spolu prikazana je pomoću stupčastog grafa na Slici 3.



Slika 3. Distribucija pacijenata po spolu

Pogledali smo još kakva je distribucija pacijenata po indexu tjelesne mase. Distribuciju vidimo u obliku stupčastog grafa na Slici 4.



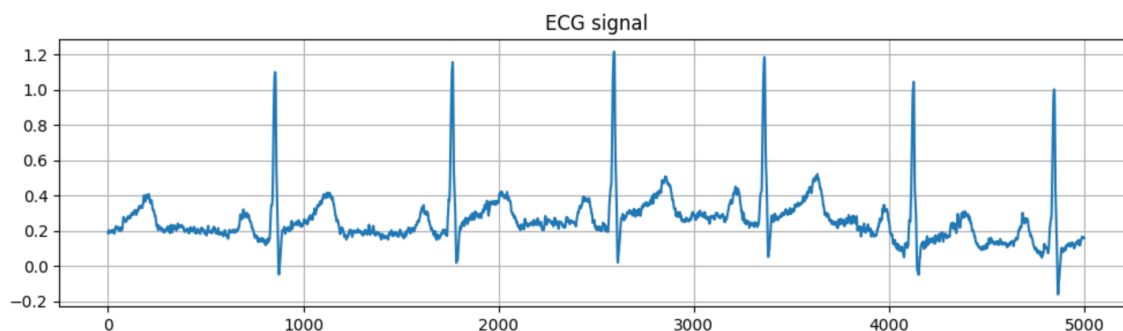
Slika 4. Distribucija indeksa tjelesne mase pacijenata

Na temelju analize, uočili smo da prosječna dob pacijenata u našem skupu podataka iznosi 32.6 godina, s prilično širokim rasponom, s obzirom na to da je standardna devijacija 14.89. Ovo ukazuje na raznolikost u dobnim skupinama, ali s naglaskom na mlađu populaciju. Što

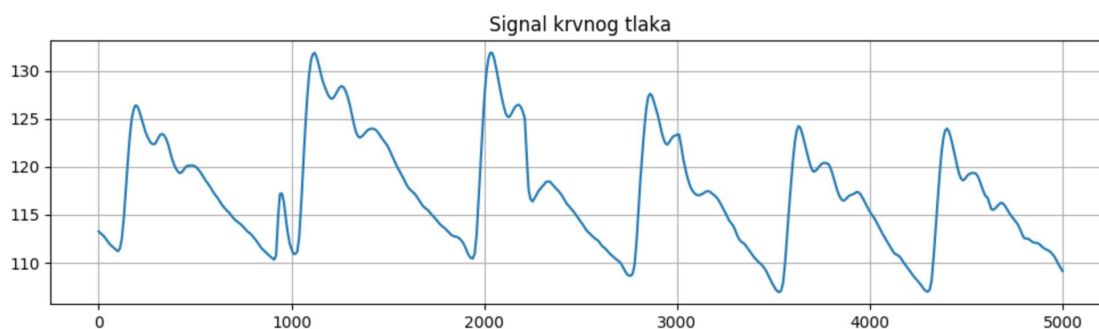
se tiče indeksa tjelesne mase (BMI), prosječna vrijednost je 23.7, s standardnom devijacijom od 4.45. Ove brojke sugeriraju da većina pacijenata ima zdravu tjelesnu masu, s obzirom na to da se preporučene vrijednosti BMI-a kreću u rasponu od 18.5 do 24.9, prema smjernicama Nacionalne zdravstvene službe Ujedinjenog Kraljevstva (NHS) [13]. Ovi podaci nam pružaju uvid u to da je naša istraživana skupina sastavljena uglavnom od mladih i zdravih pojedinaca, koji vjerojatno vode aktivan životni stil. Međutim, važno je napomenuti da iako prosječni BMI upućuje na zdravu tjelesnu masu, raspon vrijednosti naglašava varijabilnost unutar skupine.

2.2. Prikaz i ocjena kvalitete signala

U bazi podataka se nalazi par signala za svakog pacijenta, signal EKG-a i krvnog tlaka jednake duljine koji su mjereni u isto doba. Možemo vidjet na primjeru prvog pacijenta kako izgleda prvih 5000 točaka mjerenja EKG signala na Slici 5 te kako izgleda krvni tlak na Slici 6.



Slika 5, Vizualizacija tipičnog EKG signala



Slika 6, Vizualizacija signala krvnog tlaka

Neki signali su izbačeni iz skupa podataka i postoje dva kriterija prema kojima smo odlučili signal uzeti ili odbaciti.

1. EKG signal ne zadovoljava osnovne kriterije.
2. Signali krvnog tlaka izlaze iz okvira mjerenja.

2.2.1. EKG signal ne zadovoljava osnovne kriterije

Za implementaciju značajki su se koristile gotove funkcije iz python biblioteke heartpy. Funkcija koje se koristila je heartpy.process [7], kada signal nije dovoljno dobre kvalitete funkcija vraća upozorenje prikazano na Slici 7.

```
BadSignalWarning:
-----
Could not determine best fit for given signal. Please check the source signal.
Probable causes:
- detected heart rate falls outside of bpmmin<->bpmmax constraints
- no detectable heart rate present in signal
- very noisy signal (consider filtering and scaling)
If you're sure the signal contains heartrate data, consider filtering and/or scaling first.
-----
```

Slika 7, Upozorenje lošeg signala

2.2.2. Signal krvnog tlaka izlazi iz okvira mjerenja

Signal krvnog tlaka je dolazio u obliku kontinuiranog zapisa s frekvencijom uzorkovanja $f_s = 1000$ Hz. Da bi uspješno kategorizirali krvni tlak trebali smo za svaki signal izračunati vrijednost dijastoličkog i sistoličkog krvnog tlaka. Izračun sistoličkog i dijastoličkog krvnog tlaka iz signala krvnog tlaka provodi se kroz metodologiju segmentacije signala. Signal krvnog tlaka se prvo podijeli na pojedinačne segmente. Za svaki segment, identificiraju se maksimalne i minimalne vrijednosti, koje reprezentiraju vrhove i doline tlaka unutar svakog ciklusa. Nakon toga, provodi se izračunavanje prosječnih vrijednosti svih maksimalnih vrijednosti (vrhovi), što predstavlja sistolički krvni tlak, i prosječnih vrijednosti svih minimalnih vrijednosti (doline), što odgovara dijastoličkom krvnom tlaku. Nakon što smo odredili vrijednosti sistoličkog i dijastoličkog krvnog tlaka provjerili smo dali vrijednosti postižu nemoguće vrijednosti. Kriterij odbacivanja je bio da je izmjeren dijastolički tlak manji od 10 mmHg.

2.3. Finalni set podataka

Detektirano je 69 signala koji ne zadovoljavaju prethodno definirane kriterije. S obzirom na to da čini svega 6% ukupnog skupa podataka odlučili smo te signale odbaciti. Od početnih 1121 zapisa skup podataka sada sadrži 1052 zapisa.

3. Pred obrada signala

Prije ekstrakcije značajki za model strojnog učenja, potrebno je napraviti pred procesiranje signala EKG-a. Postupak je rađen u dva koraka:

- Smanjivanje buke i micanje bazne linije
- Normalizacija signala

Nakon ovog procesa imat ćemo signale iz kojih možemo izračunati kvalitetne značajke i napraviti Glavnu tablicu za učenje modela strojnog učenja.

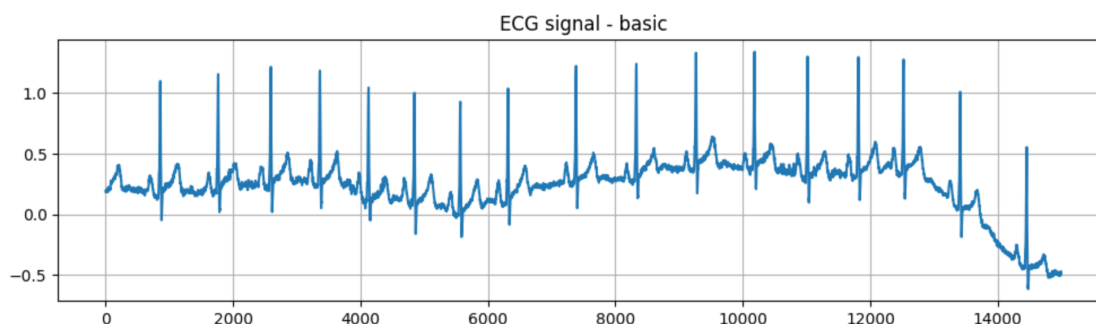
3.1. Filtriranje EKG signala i micanje osnovne linije signala

Raspon frekvencija koji se smatra relevantnim za informacije iz EKG-a kreće se od 0,05 do 100 Hz prema referenci [13]. Međutim, spektar signala EKG-a se može ograničiti do 50 Hz. Vjeruje se da se time zadržavaju sve bitne informacije iz EKG-a, kako je navedeno u izvoru [14]. S obzirom na to primijenili smo Butterworthov polu pojasni filter s donjom graničnom frekvencijom 0,5 Hz i gornjom graničnom frekvencijom 50 Hz. Postavljena donja granica od 0,5 Hz odabrana je kako bi se uklonila osnovna linija (baseline) signala [15].

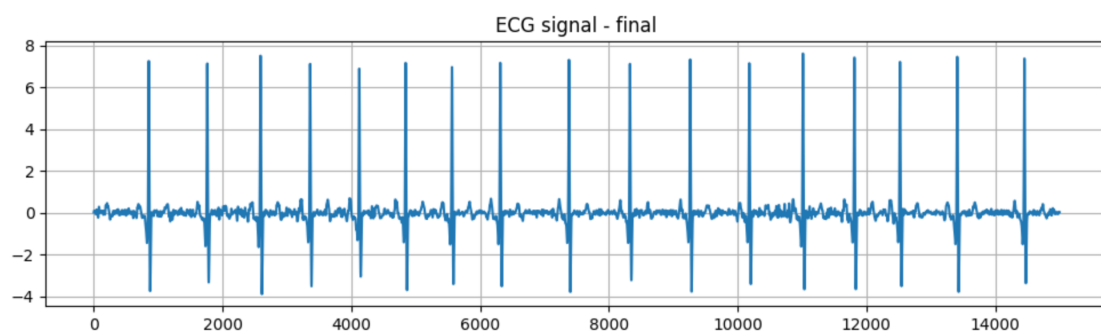
3.2. Normalizacija signala

Normalizacija signala je ključni korak u obradi biomedicinskih podataka, poput EKG signala, jer osigurava da su svi podaci na ujednačenoj skali. Ovaj proces smanjuje pristranost uzrokovanu varijacijama u amplitudi signala, poboljšava učinkovitost algoritama strojnog učenja i statističkih analiza te smanjuje utjecaj šuma i artefakata. Kroz normalizaciju, analitičari mogu preciznije i pouzdanije uspoređivati i analizirati podatke, što je posebno važno u medicinskim istraživanjima i kliničkoj praksi.

Na Slici 8 vidimo sirovi signal u 15,000 točaka te na Slici 9 vidimo signal nakon pred obrade.



Slika 8. Vizualizacija sirovog signala



Slika 9. Vizualizacija obrađenog signala

4. Odabir značajki signala

Pri odabiru značajki za algoritme strojnog učenja u kontekstu analize biomedicinskih signala, kao što su EKG i slični podaci, važno je odabrati one značajke koje najbolje predstavljaju fiziološke i statističke aspekte signala. Svaka odabrana značajka donosi jedinstveni uvid u karakteristike signala i ponašanje pacijenta.

1. Pokretljivost signala [4]: Mjeri koliko je signal promjenjiv ili "pokretan". U kontekstu EKG-a, to može ukazivati na varijabilnost srčanog ritma, što je važno za identificiranje aritmija. Za danu točku signala $x_i, i = 1, \dots, N$ gdje je N duljina signala računamo varijaciju prvog reda $d_j = x_{j+1} - x_j$. Zatim računamo faktore prvog reda S_0 i S_1 :

$$S_0 = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N}}$$
$$S_1 = \sqrt{\frac{\sum_{j=2}^{N-1} d_j^2}{N-1}}$$

Zatim računamo pokretljivost signala kao omjer faktora prvog reda S_1 i S_0

$$\text{Pokretljivost signala} = \frac{S_1}{S_0}$$

2. Kompleksnost signala [4]: Ukazuje na raznolikost njegovih oscilacija. Visoka kompleksnost može ukazivati na neregularnost u srčanom ritmu, što može biti znak patoloških stanja.

$$S_2 = \sqrt{\frac{\sum_{k=3}^{N-2} g_k^2}{N-2}}$$

Gdje je varijacija drugog reda signala predstavljena sa:

$$g_k = d_{k+1} - d_k$$

3. Fraktalna dimenzija [4]: Ova značajka se koristi za kvantificiranje kompleksnosti signala na način koji prelazi uobičajene linearne mjere. Može dati uvid u složene obrasce srčanog ritma koji su tipični za određena medicinska stanja.

Značajka je računata prema Algoritmu Higuchija [16].

4. Entropija [4]: Mjeri neurednost ili neizvjesnost signala. U medicinskim aplikacijama, visoka entropija može ukazivati na nepravilnosti u srčanom ritmu.

$$\text{Entropija} = \sum_{i=0}^{N-1} p_i \log\left(\frac{1}{p_i}\right)$$

Gdje p_i predstavlja vjerojatnost svakog ishoda x_i unutar EKG signala.

5. Otkucaji u minuti (BPM) [12]: Ovo je osnovna mjera srčane frekvencije i od vitalnog je značaja za svaku analizu srčanog zdravlja.

$$BPM = \frac{60}{IBI}$$

6. Interval između otkucaja (IBI) [12]: Razmak između otkucaja srca, koristan za analizu varijabilnosti srčanog ritma.

$$IBI = \text{Srednja Vrijednost}(t_{i+1} - t_i)$$

Gdje su t_i vremenske oznake svakog otkucaja srca.

7. Stopa disanja [17] [12]: Često se koristi u kombinaciji s EKG podacima za pružanje uvida u autonomni živčani sustav.
8. SDNN (Standardna Devijacija NN Intervala) [12]: Mjera varijabilnosti srčanog ritma, korisna za identificiranje stresnih stanja ili kardiovaskularnih problema.
9. RMSSD (Kvadratni Korijen Srednje Vrijednosti Uzastopnih Razlika) [12]: Još jedna mjera varijabilnosti srčanog ritma, usredotočena na kratkoročne varijacije.
10. PNN50 (Postotak Uzastopnih NN Intervala s Razlikom Većom od 50 ms) [12]: Mjera dugoročne varijabilnosti srčanog ritma.
11. Dob [4]: Godine mogu bit važan faktor u analizi zdravlja srca.
12. Spol: Može utjecati na srčane obrasce i rizike.
13. Indeks tjelesne mase (BMI): BMI je važan čimbenik u procjeni općeg zdravlja i može utjecati na srčano zdravlje.

Implementacija značajki: Pokretljivost signala, Kompleksnost signala, Fraktalna dimenzija i Entropije izvedena je kao novo definirana funkcija koju vidimo u Programskom kodu -1. Dok sve ostale značajke su implementirane kroz gotovu funkciju Process() iz python biblioteke Heartpy [12] koja je prikazana na Slici - 10.

```

def calculate_ecg_features(ecg_signal):
    N = len(ecg_signal)

    # Signal Mobility
    S0 = np.sqrt(np.sum(np.square(ecg_signal)) / N)
    dj = np.diff(ecg_signal)
    S1 = np.sqrt(np.sum(np.square(dj)) / (N - 1))
    mobility = S1 / S0

    # Signal Complexity
    gk = np.diff(dj)
    S2 = np.sqrt(np.sum(np.square(gk)) / (N - 2))

    # Fractal Dimension
    kmax = 10 # Choose the maximum value of k
    lm_list = []
    for k in range(1, kmax + 1):
        num_segments = int(np.floor((N - 1) / k))
        lm = 0.0
        for m in range(1, k + 1):
            subseries = ecg_signal[m-1:N-m+1:k]
            lm += np.sum(np.abs(np.diff(subseries))) * (N -
1) / (num_segments * k)
        lm /= k
        lm_list.append(lm)
    k_values = 1 / np.arange(1, kmax + 1)
    fractal_dimension = np.polyfit(np.log(k_values),
np.log(lm_list), 1)[0]

    #Entropy
    prob_dist = np.histogram(ecg_signal, bins=256,
density=True)[0]

    # Filter out zero probabilities to avoid logarithm issues
    prob_dist = prob_dist[prob_dist > 0]

    # Calculate the entropy using the Shannon entropy formula
    entropy = -np.sum(prob_dist * np.log2(prob_dist))

    return mobility, S2, fractal_dimension, entropy

```

Kôd 4. Implementacija značajki signala

```

wd, m = hp.process(hp.scale_data(filtered_ecg[0]), sample_rate)
bpm=m['bpm']
ibi=m['ibi']
breathingrate=m['breathingrate']
sdnn=m['sdnn']
rmssd=m['rmssd']
pnn50=m['pnn50']
print("Otkucaji u minuti:"+str(bpm))
print("Interval između otkucaja:"+str(ibi))
print("Stopa disanja:"+str(breathingrate))
print("Standardna Devijacija NN Intervala:"+str(sdnn))
print("Kvadratni Korijen Srednje Vrijednosti Uzastopnih Razlika:"+str(rmssd))
print("Postotak Uzastopnih NN Intervala s Razlikom Većom od 50 ms:"+str(pnn50))

Otkucaji u minuti:70.40547876560738
Interval između otkucaja:852.2064056939502
Stopa disanja:0.14615914643058484
Standardna Devijacija NN Intervala:62.1674363063415
Kvadratni Korijen Srednje Vrijednosti Uzastopnih Razlika:63.81802057762081
Postotak Uzastopnih NN Intervala s Razlikom Većom od 50 ms:0.3850267379679144

```

Slika 10, Korištenje funkcije heartpy.process()

Svaka od ovih značajki pruža drugačiju perspektivu na zdravstveno stanje pacijenta, i zajedno omogućuju detaljan uvid u srčane funkcije i potencijalne zdravstvene probleme. Kombiniranjem ovih značajki, algoritmi strojnog učenja mogu preciznije analizirati i predviđati zdravstvena stanja, doprinoseći unaprjeđenju dijagnostičkih alata i personalizirane medicine.

5. Klase krvnog tlaka

Klasifikaciju krvnog tlaka smo podijelili na dva slučaja, klasifikaciju u 2 klase: sniženi i povišeni te klasifikaciju u 3 klase: sniženi, normalni i povišeni. U Tablici 3 vidimo kriterij klasifikacije za slučaj 2 klase te u Tablici 4 kriterije za 3 klase. Također možemo vidjeti koliko je primjeraka signala u svakoj od klasa. Svaku od klasa smo identificirali s oznakom ID gdje u slučaju 2 klase 0 predstavlja sniženi i 1 predstavlja povišeni dok u slučaju 3 klase 0 predstavlja sniženi, 1 predstavlja normalni i 2 predstavlja povišeni krvni tlak.

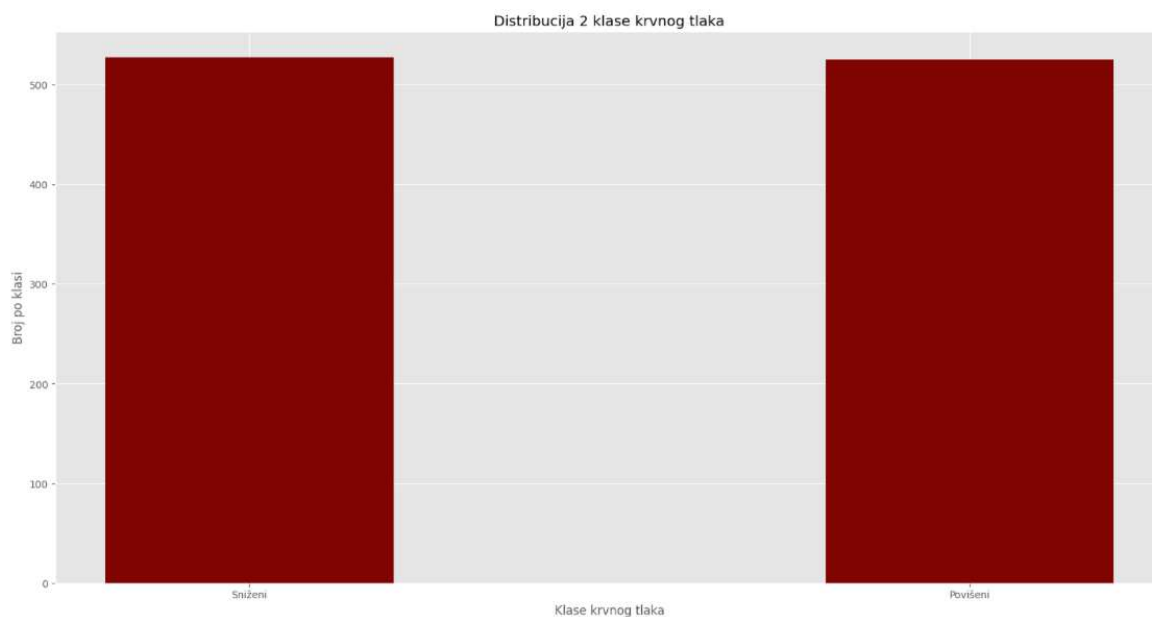
Tablica 3. Podjela na 2 klase krvnog tlaka

Category	SBP(mmHg)	DBP(mmHg)	ID	Ukupno
Sniženi	<120	<80	0	527
Povišeni	>=120	>=80	1	525

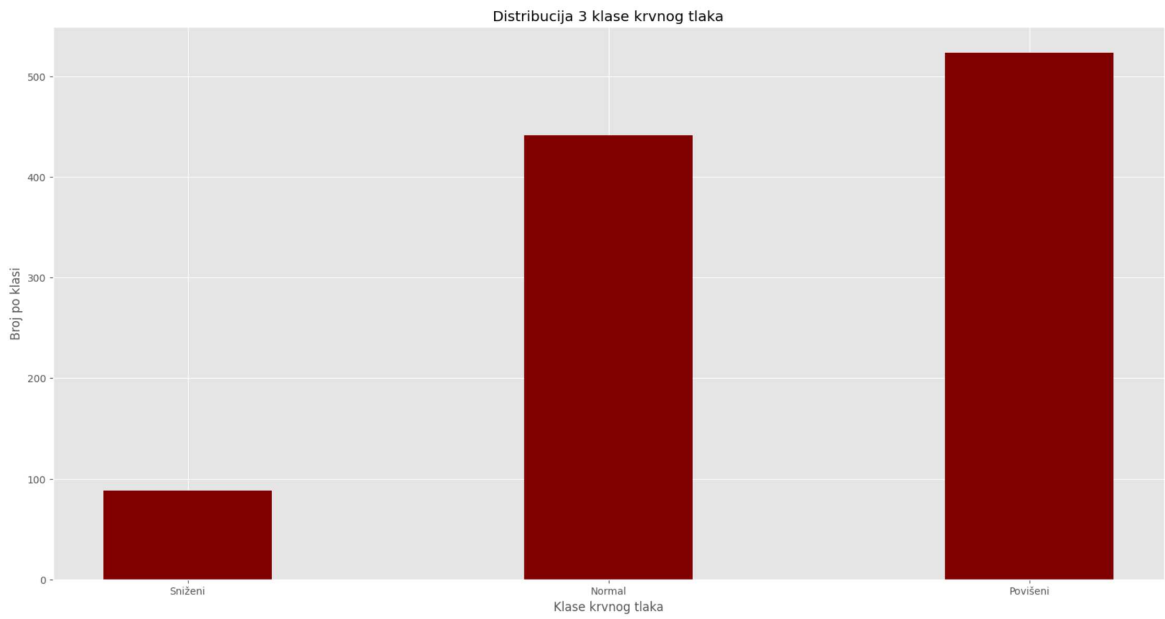
Tablica 4. Podjela na 3 klase krvnog tlaka

Category	SBP(mmHg)	DBP(mmHg)	ID	Ukupno
Sniženi	<90	<60	0	88
Normalni	90-120	60-80	1	441
Povišeni	>=120	>=80	2	523

Distribuciju klasa redom vidimo u obliku stupčastog grafa na Slici - 11 za 2 klase i Slici – 12 za 3 klase.



Slika 11. Distribucija krvnog tlaka na 2 klase



Slika 12- Distribucija krvnog tlaka na 3 klase

Vidimo da je podjela na 2 klase podjednaka dok u slučaju 3 klase sniženi krvni tlak predstavlja 8.4% podataka, normalni predstavlja 41.9% podataka i povišeni 49.7% podataka.

6. Izrada glavne tablice

Nakon što su svi signali pred procesirani i odabrane značajke bilo je potrebno izraditi glavnu tablicu kao ulazni parametar za učenje modela strojnog učenja. Za svaki od 1052 signala su se računale značajke te formirale vektor značajki. Nakon što je vektor značajki bio formiran on se dodao u listu X koja predstavlja glavnu tablicu. Implementaciju procesa vidimo u programskom Kôdu 2.

```
from scipy.stats import skew, kurtosis
features=[]
for i in range(len(filtered_ecg)):
    wd, m = hp.process(hp.scale_data(filtered_ecg[i]),
sample_rate)
    mobility, s2, frac, entropy=
calculate_ecg_features(filtered_ecg[i])
    bpm=m['bpm']
    ibi=m['ibi']
    breathingrate=m['breathingrate']
    sdnm=m['sdnm']
    rmssd=m['rmssd']
    pnn50=m['pnn50']
    if(np.isnan(mobility) or np.isnan(s2) or np.isnan(frac)):
        mobility, s2, frac, auto, entropy=(0,0,0,0,0)
    age=filtered_age[i]
    sex=filtered_sex[i]
    bmi=filtered_bmi[i]
    feature_line=[bpm, ibi, breathingrate, sdnm, rmssd,
pnn50, mobility, s2, frac, entropy, age, sex, bmi]
    features.append(feature_line)

X=np.vstack(features)
```

Kôd 2. Izrada glavne tablice

Uz glavnu tablicu formirani su vektori klasa odnosno ciljani stupac za modele strojnog učenja. Vektore smo označili oznakom y kada smo trenirali modele.

6.1. Podijele na skup za treniranje i testiranje

Glavnu tablicu smo podijelili na skup za učenje i skup za testiranje modela strojnog učenja. Implementacije podjele je kroz funkciju `train_test_split()` iz Python biblioteke `sklearn`. Skup za učenje čini 80% podataka, a skup za testiranje ostalih 20%. Razina slučajnosti uzeta je 42 što predstavlja u funkciji kojim redoslijedom će podijeliti podatke kako bi svaki put kada iznova pokrenemo programski kod bio reproduciran isti rezultat. Nakon takve podjele u Tablici 5 vidimo podjelu skupa za učenje i skupa za treniranje po klasama za slučaj 2 klase i u Tablici 6 za slučaj 3 klase.

Tablica 5. Razdioba klasa u skupovima za treniranje i testiranje u slučaju 2 klase

Klasa	0	1
Treniranje	420	421
Testiranje	107	104

Tablica 6. Razdioba klasa u skupovima za treniranje i testiranje u slučaju 3 klase

Klasa	0	1	2
Treniranje	68	351	422
Testiranje	20	90	101

7. Izrada klasifikacijskog modela strojnog učenja

Kada se suočavamo s izazovom klasifikacije u strojnom učenju, izbor pravog modela je ključan. Stablo odlučivanja, Slučajna šuma, Potporni vektorski stroj i Logistička regresija predstavljaju četiri različita, ali iznimno učinkovita pristupa ovom zadatku. Svaki od ovih modela donosi jedinstvene prednosti i pristupe rješavanju problema klasifikacije. Posebno se ističe u određenim scenarijima klasifikacije. Stabla odlučivanja i Slučajna šuma su vrlo interpretirajući, dok SVM i Logistička regresija nude snažne performanse u složenim i visoko-dimenzionalnim prostorima.

7.1. Stablo odlučivanja

Stabla odlučivanja su temeljna alatka u strojnom učenju zbog njihove jednostavnosti i intuitivnosti. Ovaj model koristi seriju binarnih pravila (da/ne) za donošenje odluka, što ga čini lako razumljivim čak i osobama bez tehničkog znanja. U srcu stabla odlučivanja je hijerarhijska struktura koja na svakom čvoru postavlja pitanje o jednoj od značajki, vodeći do konačne odluke na listovima stabla. Ovo omogućuje ne samo klasifikaciju, već i dublje razumijevanje faktora koji utječu na predikciju. Međutim, stabla odlučivanja su sklonija prekomjernom uklapanju, posebno u slučajevima sa složenim skupovima podataka.

Implementaciju stabla odlučivanja vidimo u programskom Kôdu 3.

```
# Decision Tree with Hyperparameter Tuning
dt_param_grid = {'max_depth': [None, 10, 20, 30],
                 'min_samples_split': [2, 5, 10]}
dt_classifier = GridSearchCV(DecisionTreeClassifier(),
                             dt_param_grid, cv=5)
dt_classifier.fit(X_train, y_train)
dt_pred = dt_classifier.predict(X_test)
```

Kôd 3. Program za implementaciju stabla odlučivanja

7.2. Slučajna šuma

Slučajna šuma, s druge strane, predstavlja ansambl metodu koja koristi više stabala odlučivanja da poboljša točnost i prevlada ograničenja pojedinačnih stabala. Kroz proces slučajnog uzorkovanja podataka i značajki za svako stablo, ovaj model smanjuje varijabilnost i povećava točnost predviđanja. Rezultat je robusniji model koji je otporniji na prekomjerno uklapanje i bolje generalizira na neviđenim podacima. Svako stablo donosi nezavisnu predikciju, a konačna odluka je rezultat agregacije ovih predikcija, često putem većinske glasovne sheme.

Implementaciju slučajne šume vidimo u programskom Kôdu 4.

```
# Random Forest with Hyperparameter Tuning
rf_param_grid = {'n_estimators': [50, 100, 200], 'max_depth':
[None, 10, 20, 30], 'min_samples_split': [2, 5, 10]}
rf_classifier =
GridSearchCV(RandomForestClassifier(class_weight=class_weight
_dict), rf_param_grid, cv=5)
rf_classifier.fit(X_train, y_train)
rf_pred = rf_classifier.predict(X_test)
```

Kôd 4. Program za implementaciju slučajne šume

7.3. Potporni vektorski stroj (SVM)

SVM [20] je jedan od najmoćnijih i najfleksibilnijih modela u strojnom učenju, posebno kada se radi o visoko-dimenzionalnim i kompleksnim datasetovima. Ovaj model pronalazi hiperplohu ili set hiperploha u visoko-dimenzionalnom prostoru koja najbolje razdvaja različite klase. Ključna karakteristika SVM-a je njegova sposobnost maksimiziranja margine između različitih klasa, što rezultira visokom generalizacijskom sposobnošću. Još jedna prednost SVM-a je njegova sposobnost korištenja kernel trikova, omogućavajući modelu da efikasno rješava i ne-linearne probleme klasifikacije.

Implementaciju potpornog vektorskog stroja vidimo u programskom Kôdu 5.

```
# Support Vector Machine (SVM) with Hyperparameter Tuning
svm_param_grid = {'C': [0.1, 1, 10], 'kernel': ['linear',
'rbf']}
svm_classifier =
GridSearchCV(SVC(class_weight=class_weight_dict,
probability=True), svm_param_grid, cv=5)
svm_classifier.fit(X_train, y_train)
svm_pred = svm_classifier.predict(X_test)
```

Kôd 5. Program za implementaciju potpornog vektorskog stroja (SVM)

7.4. Logistička regresija

Logistička regresija [20] je probabilistički diskriminativni model. Unatoč nazivu, nije riječ o regresiji nego o klasifikaciji. Koristi se za procjenu vjerojatnosti da određeni uzorak pripada jednoj od dvije moguće kategorije. Ovaj model je posebno koristan zbog svoje jednostavnosti i efikasnosti u pružanju probabilističkih ishoda za klasifikaciju. Logistička regresija koristi logističku funkciju za pretvaranje linearnih kombinacija ulaznih značajki u vjerojatnosti, nudeći jasan i jednostavan način za modeliranje binarnih ishoda.

Implementaciju logističke regresije vidimo u programskom Kôdu 6.

```
# Logistic Regression with Hyperparameter Tuning
lr_param_grid = {'C': [0.1, 1, 10]}
lr_classifier =
GridSearchCV(LogisticRegression(class_weight=class_weight_dic
t), lr_param_grid, cv=5)
lr_classifier.fit(X_train, y_train)
lr_pred = lr_classifier.predict(X_test)
```

Kôd 6. Program za implementaciju logističke regresije

Proces se sastoji od nekoliko ključnih koraka: standardizacije, balansiranja klasa, hiperparametarskog podešavanja, treniranja modela i evaluacije modela.

7.5. Standardizacija numeričkih značajki

Za modele poput SVM-a, važno je standardizirati značajke tako da imaju slične raspona. To se postiže korištenjem funkcije StandardScaler, koja normalizira značajke oduzimanjem srednje vrijednosti i dijeljenjem sa standardnom devijacijom.

7.6. Balansiranje klasa

Kada se suočavamo s nebalansiranim klasama u setu podataka za strojno učenje, postoji značajan rizik da model razvije pristranost prema dominantnijoj klasi. Ovo je zbog činjenice da modeli strojnog učenja, posebno oni temeljeni na metrikama poput točnosti, mogu biti obmanuti brojčanom premoću jedne klase nad drugom. Kao rezultat, model može postati bolji u predviđanju dominantne klase, dok zanemaruje manje zastupljene klase, što može dovesti do neadekvatnih i nepouzdanih predviđanja.

Da bi se ovaj problem riješio, koristi se mehanizam prilagodbe težina klasa, poznat kao `class_weight`. Ovaj pristup uključuje dodjeljivanje većih težina manje zastupljenim klasama i manjih težina dominantnijim klasama. Cilj je osigurati da model ne ignorira manje zastupljene klase i da se postigne bolja ravnoteža u točnosti predviđanja među svim klasama.

7.7. Optimizacija hiperparametara i treniranje modela

Optimizacija hiperparametara [20] je ključni korak u procesu izrade modela strojnog učenja, gdje se odabiru optimalne vrijednosti hiperparametara koji kontroliraju proces učenja modela. Svaki model - Stablo odlučivanja, Slučajna šuma, SVM i Logistička regresija - podvrgnut je hiperparametarskom podešavanju koristeći GridSearchCV.

Hiperparametri su konfiguracijske varijable koje se koriste za kontrolu procesa učenja modela strojnog učenja. Za razliku od parametara modela, koji se uče iz podataka tijekom procesa učenja, hiperparametri se postavljaju prije početka učenja i ostaju nepromijenjeni tijekom tog procesa.

Grid Search je tehnika za sistematsko ispitivanje više kombinacija hiperparametara. GridSearchCV provodi ovaj proces tako da prolazi kroz sve moguće kombinacije hiperparametara definirane u rešetki (grid). Za svaku kombinaciju hiperparametara, model

se trenira i evaluira koristeći unakrsnu validaciju (cross-validation), koja osigurava da se svaka kombinacija testira na različitim pod skupovima podataka. Na kraju, Grid Search odabire kombinaciju hiperparametara koja daje najbolje performanse, prema zadanim mjerama uspješnosti.

Za svaki model definiraj je skup hiperparametara koji se trebaju isprobati:

- Stablo odlučivanja (Decision Tree): Istražuju se različite kombinacije maksimalne dubine stabla (`max_depth`) i minimalnog broja uzoraka potrebnog za podjelu čvora (`min_samples_split`).
- Slučajna šuma (Random Forest): Slično kao i za stablo odlučivanja, ali dodatno uključuje broj stabala u šumi (`n_estimators`).
- Potporni vektorski stroj (SVM): Istražuje različite vrijednosti parametra regularizacije (`C`) i tipove jezgara (kernel).
- Logistička regresija: Fokusira se na parametar regularizacije (`C`).

Kroz ovaj proces, Grid Search omogućava pronalazak optimalnih hiperparametara za svaki model, što dovodi do boljih performansi modela i pouzdanijih predviđanja.

7.8. Evaluacija modela

Nakon treniranja, modeli se testiraju na testnom skupu. Evaluacija uključuje izračunavanje točnosti, preciznosti i odziva za svaku kategoriju, što daje uvid u performanse modela u klasifikaciji različitih klasa.

Točnost [21] je najjednostavnija mjera i predstavlja postotak ispravno klasificiranih primjera. Izračunava se kao omjer broja ispravno klasificiranih primjera prema ukupnom broju primjera

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

Preciznost [21] mjeri koliko su predikcije modela točne kada model predviđa određenu klasu. Drugim riječima, to je omjer ispravno pozitivnih predviđanja prema ukupnom broju pozitivnih predviđanja.

$$\text{Preciznost} = \frac{TP}{TP + FP}$$

Odziv [21] mjeri koliko je model sposoban identificirati sve relevantne slučajeve određene klase. To je omjer ispravno pozitivnih predviđanja prema ukupnom broju stvarnih pozitivnih slučajeva.

$$\text{Odziv} = \frac{TP}{TP + FN}$$

Gdje je TP pravo pozitivno, TN pravo negativno, FP lažno pozitivno i FN lažno negativno.

8. Rezultati

8.1. Klasifikacijski modeli 2 klase

Najbolje rezultate za klasifikacijski model u 2 klase smo dobili za set hiperparametara koje vidimo u Tablici 7.

Tablica 7. Hiperparametri za model 2 klase

Model	Hiperparametar 1	Hiperparametar 2	Hiperparametar 3
Slučajna šuma	Max_depth:20	Min_samples_split:2	N_estimators:200
Stablo odlučivanja	Max_depth:10	Min_samples_split:10	/
SVM	C:1	Kernel:“Linear“	/
Logistička Regresija	C:10	/	/

Na testnom skupu od 211 pacijenata konfuzijsku matricu za model Stabla odlučivanja vidimo u Tablici 8.

Tablica 8. Konfuzijska matrica Stabla odlučivanja

Stablo odlučivanja	0	1
0	58	49
1	49	55

Prema čemu računamo rezultate točnosti, preciznosti i odziva vidljivih u Tablici 9.

Tablica 9. Rezultati modela Stabla odlučivanja

Mjera	Rezultat
Točnost	53.55%
Preciznost	54.00%
Odziv	54.00%

Konfuzijsku matricu za model Slučajne šume vidimo u Tablici 10

Tablica 10. Konfuzijska matrica Slučajne šume

Stablo odlučivanja	0	1
0	75	32
1	48	56

Prema čemu računamo rezultate točnosti, preciznosti i odziva vidljivih u Tablici 11.

Tablica 11. Rezultati modela Slučajne šume

Mjera	Rezultat
Točnost	62.09%
Preciznost	62.00%
Odziv	62.00%

Konfuzijsku matricu za model SVM-a vidimo u Tablici 12.

Tablica 12. Konfuzijska matrica SVM-a

Stablo odlučivanja	0	1
0	49	58
1	30	74

Prema čemu računamo rezultate točnosti, preciznosti i odziva vidljivih u Tablici 13.

Tablica 13. Rezultati modela SVM-a

Mjera	Rezultat
Točnost	58.29%
Preciznost	59.00%
Odziv	57.99%

Konfuzijsku matricu za model Logističke regresije vidimo u Tablici 14.

Tablica 14. Konfuzijska matrica Logističke regresije

Stablo odlučivanja	0	1
0	57	50
1	32	72

Prema čemu računamo rezultate točnosti, preciznosti i odziva vidljivih u Tablici 15.

Tablica 15. Rezultati modela Logističke regresije

Mjera	Rezultat
Točnost	61.14%
Preciznost	62.00%
Odziv	61.00%

8.2. Klasifikacijski modeli 3 klase

Najbolje rezultate za klasifikacijski model u 3 klase smo dobili za set hiperparametara koje vidimo u Tablici 16.

Tablica 16. Hiperparametri za model 3 klase

Model	Hiperparametar 1	Hiperparametar 2	Hiperparametar 3
Slučajna šuma	Max_depth:30	Min_samples_split:2	N_estimators:50
Stablo odlučivanja	Max_depth:10	Min_samples_split:10	/
SVM	C:10	Kernel:“rbf“	/
Logistička Regresija	C:10	/	/

Na testnom skupu od 211 pacijenata konfuzijsku matricu za model Stabla odlučivanja vidimo u Tablici 17

Tablica 17. Konfuzijska matrica Stabla odlučivanja

Stablo odlučivanja	0	1	2
0	2	10	8
1	7	40	43
2	11	33	57

Prema čemu računamo rezultate točnosti, preciznosti i odziva vidljivih u Tablici 18.

Tablica 18. Rezultati modela Stabla odlučivanja

Mjera	Rezultat
Točnost	46.92%
Preciznost	47.00%
Odziv	47.00%

Konfuzijsku matricu za model Slučajne šume vidimo u Tablici 19.

Tablica 19. Konfuzijska matrica Slučajne šume

Stablo odlučivanja	0	1	2
0	0	14	6
1	3	64	23
2	0	44	57

Prema čemu računamo rezultate točnosti, preciznosti i odziva vidljivih u Tablici 20.

Tablica 20. Rezultati modela Slučajne šume

Mjera	Rezultat
Točnost	57.35%
Preciznost	54.00%
Odziv	56.99%

Konfuzijsku matricu za model SVM-a vidimo u Tablici 21.

Tablica 21. Konfuzijska matrica SVM-a

Stablo odlučivanja	0	1	2
0	2	10	8
1	26	40	24
2	13	32	56

Prema čemu računamo rezultate točnosti, preciznosti i odziva vidljivih u Tablici 22.

Tablica 22. Rezultati modela SVM-a

Mjera	Rezultat
Točnost	46.45%
Preciznost	52.00%
Odziv	46.00%

Konfuzijsku matricu za model Logističke regresije vidimo u Tablici 23.

Tablica 23. Konfuzijska matrica Logističke regresije

Stablo odlučivanja	0	1	2
0	8	6	6
1	39	26	25
2	34	23	44

Prema čemu računamo rezultate točnosti, preciznosti i odziva vidljivih u Tablici 24.

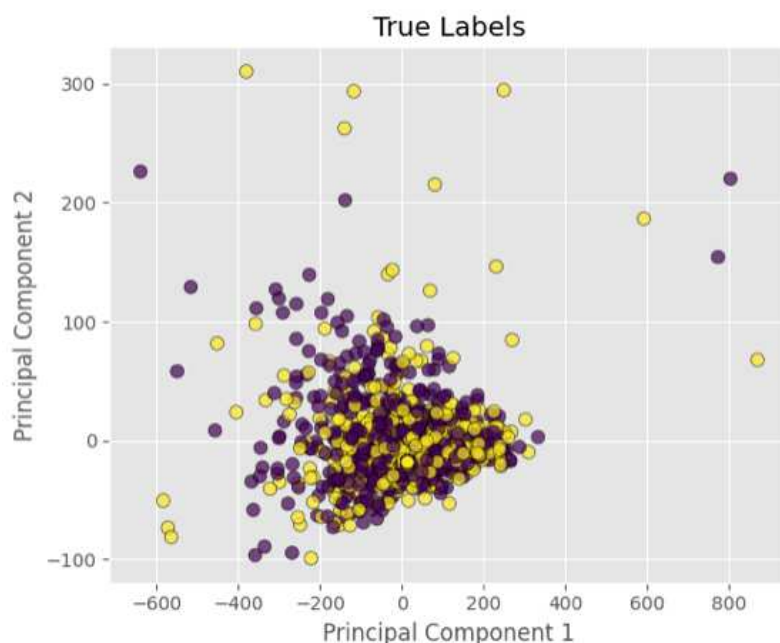
Tablica 24. Rezultati modela Logističke regresije

Mjera	Rezultat
Točnost	36.97%
Preciznost	49.00%
Odziv	37.00%

8.3. PCA analiza značajki

S obzirom na rezultate u slučaju 2 klase te uz prethodnu analizu demografskog uzorak pacijenata navedeni smo na razmišljanje dali postoji razlog zašto se podaci ne klasificiraju boljom točnošću. Iz tog razloga odlučili smo provesti PCA analizu značajki kako bi spustili dimenzionalnost u dvije dimenzije kako bi mogli bolje vizualizirati i razumjeti podatke. Analiza glavnih komponenti (PCA) [22] je statistička tehnika koja se koristi u području strojnog učenja i analize podataka za smanjenje dimenzionalnosti podataka, istovremeno zadržavajući što je više moguće varijance. Osnovna ideja PCA je transformirati velik broj mogućih međusobno zavisnih varijabli u manji broj nezavisnih varijabli, poznatih kao glavne komponente. Ovaj proces omogućuje bolje razumijevanje strukture i distribucije podataka.

Implementacija je odrađena s gotovim funkcijama iz Python biblioteke sklearn. Na Slici 8-1 vidimo rezultate PCA kada sve značajke transformiramo u dvije glavne komponente. Uočavamo da preklapanje glavnih komponenata izvedenih iz PCA analize ukazuje na značajnu sličnost među različitim skupinama u skupu podataka. Ova sličnost među komponentama sugerira da su razlike između klasa koje smo analizirali prilično suptilne ili možda čak neizražene u trenutno korištenim značajkama. Kao rezultat, to nam govori da modeli strojnog učenja bi mogli imati ograničen kapacitet za preciznu klasifikaciju krvnog tlaka na temelju EKG signala, što je posebno izazovno u situacijama s nejednako raspoređenim klasama.



Slika 13. PCA analiza značajki

Zaključak

Na temelju provedenih eksperimenata i analiza, možemo izvesti nekoliko ključnih zaključaka u vezi s klasifikacijom krvnog tlaka iz EKG signala koristeći različite modele strojnog učenja.

Modeli strojnog učenja pokazuju varijabilne rezultate u klasifikaciji krvnog tlaka. U scenariju s tri klase (sniženi, normalni, povišeni), slučajne šume su se pokazale kao najučinkovitiji model s točnošću od 57.35%. stablo odlučivanja i stroj potpornih vektora pokazali su se manje efikasni, s točnošću od 46.92% i 46.45% respektivno. Logistička regresija je ostvarila najnižu točnost 36.97%. U scenariju s dvije klase, model slučajnih šuma ponovno ima najveću točnost 62.09%, ali s povećanjem preciznosti i odziva u usporedbi s tro klasnim scenarijem. Zatim logistička regresija s točnošću 61.14%, koju prate stroj potpornih vektora i stablo odlučivanja s točnošću 58.29% i 53.55% respektivno.

Znatna nejednakost u distribuciji klasa u scenariju s tri klase, gdje je broj sniženih slučajeva (88) znatno manji od broja slučajeva s povišenim (523) i normalnim (441) krvnim tlakom. Ova neravnoteža može negativno utjecati na sposobnost modela da ispravno klasificira manje zastupljene klase, što je vidljivo iz relativno niskih vrijednosti preciznosti i odziva za neke modele. Dok je u slučaju dvije klase distribucija podataka podjednaka što nas je navelo da provedemo PCA analizu značajki. PCA analiza ukazuje na značajno preklapanje među klasama, što dodatno komplicira klasifikaciju. Preklapanje podataka sugerira da su razlike između klasa suptilne i možda nisu dovoljno izražene u korištenim značajkama EKG signala, što može dovesti do poteškoća u učinkovitom razlikovanju među klasama.

Prosječna dob pacijenata od 32.6 godina s relativno visokom standardnom devijacijom od 14.89 ukazuje na pristojnu raznolikost u dobnim skupinama. Međutim, s obzirom na to da je prosječna dob na nižoj strani spektra, postoji jasan naglasak na mlađu populaciju. Ovo može imati značajan utjecaj na rezultate modela, jer mlađi pacijenti generalno imaju manju vjerojatnost za probleme s krvnim tlakom u usporedbi sa starijom populacijom. Prosječni BMI od 23.7, sa standardnom devijacijom od 4.45, dodatno ukazuje na to da većina pacijenata spada u kategoriju zdrave tjelesne mase. Ovo je relevantno jer BMI može biti povezan s krvnim tlakom, pri čemu osobe s višim BMI-om često imaju veći rizik od hipertenzije.

Budući rad trebao bi se fokusirati na prikupljanje uravnoteženijeg skupa podataka, koji bi uključivao veći broj pojedinaca s povišenim i sniženim krvnim tlakom, te raznolikiji demografski profil. Također, istraživanje dodatnih značajki ili primjena naprednijih tehnika strojnog učenja, poput dubokog učenja u primjeru rada [5], moglo bi poboljšati točnost klasifikacije. Nadalje, provedba dodatne validacije modela na nezavisnim skupovima podataka također bi bila korisna za procjenu njihove opće primjenjivosti.

U zaključku, iako su modeli strojnog učenja pokazali određeni potencijal u klasifikaciji krvnog tlaka iz EKG signala, postoje značajni izazovi koji se odnose na neravnotežu u klasama, preklapanje podataka i ograničenja uzorka. Daljnji razvoj i optimizacija modela, zajedno s poboljšanjem kvalitete i raznolikosti skupa podataka, ključni su za unapređenje točnosti i pouzdanosti u klasifikaciji krvnog tlaka.

Literatura

- [1] Balažin Vučetić A., Krvni Tlak, PLIVAZDRAVLJE, (2007, ožujak). Poveznica: <https://www.plivazdravlje.hr/aktualno/clanak/16063/Krvni-tlak.html>; pristupljeno 20. siječnja 2024.
- [2] Zavod za elektroničke sustave i obradbu informacija, Elektrokardiografija, Fakultet Elektrotehnike i Računarstva, Poveznica: https://www.fer.unizg.hr/download/repository/Vjezba_1_-_Elektrokardiografija%5B1%5D.pdf; pristupljeno 20. siječnja 2024.
- [3] Sovilj S., Trogrlić D., Magjarević R. Šantić A., Biomedicinska instrumentacija, Fakultet Elektrotehnike i Računarstva, (2011). Poveznica: https://www.fer.unizg.hr/download/repository/Vjezba_05_06_-_SpO2_KrvniTlak.pdf; pristupljeno 20. siječnja 2024.
- [4] Simjanoska M, Gjoreski M, Gams M, Madevska Bogdanova A. Non-Invasive Blood Pressure Estimation from ECG Using Machine Learning Techniques. Sensors (Basel). 2018
- [5] Fuadah YN, Lim KM. Classification of Blood Pressure Levels Based on Photoplethysmogram and Electrocardiogram Signals with a Concatenated Convolutional Neural Network. Diagnostics (Basel). 2022
- [6] Sharma, M., Rajput, J.S., Tan, R.S., Acharya, U.R. Automated Detection of Hypertension Using Physiological Signals: A Review. Int. J. Environ. Res. Public Health. 2021
- [7] Schumann A., Bär K., Autonomic Aging: A dataset to quantify changes of cardiovascular autonomic function during healthy aging, (2021). Poveznica: <https://physionet.org/content/autonomic-aging-cardiovascular/1.0.0/>
- [8] Open Data Commons Open Database License v1.0 Poveznica: <https://physionet.org/content/autonomic-aging-cardiovascular/view-license/1.0.0/#> ; pristupljeno 20. siječnja 2024
- [9] ARX – Data anonymization tool (2016). Poveznica: <https://arx.deidentifier.org/>; pristupljeno 20. siječnja 2024
- [10] Biopac System MP150 System (2016). Poveznica: <https://www.biopac.com/wp-content/uploads/MP150-Systems.pdf>; pristupljeno 22. siječnja 2024
- [11] CNSystems Medizintechnik GmbH Task Force monitor. Poveznica: <https://www.cnsystems.com/products/task-force-monitor/>; pristupljeno 22. siječnja 2024
- [12] Mihaljević S, Fotopletizmografsko ispitivanje simpatičke blokade izazvane lumbalnom epiduralnom anestezijom bupivakainom i S-(+) ketaminom. Doktorska disertacija, Sveučilište u Zagrebu.(2006)
- [13] Nacionalna zdravstvena služba Ujedinjenog Kraljevstva (NHS). Poveznica: <https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/>; pristupljeno 24. siječnja 2024

- [14] Python Heart Rate Analysis Toolkit, (2023). Poveznica: <https://python-heart-rate-analysis-toolkit.readthedocs.io/en/latest/heartpy.heartpy.html>;
- [15] Rangayyan R.M. Biomedical Signal Analysis. Volume 33 John Wiley & Sons; Hoboken, NJ, USA: 2015.
- [16] Mahdiani S., Jeyhani V., Peltokangas M., Vehkaoja A. Is 50 Hz high enough ECG sampling frequency for accurate HRV analysis?; Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Milan, Italy. 25–29 August 2015
- [17] Luo Y., Hargraves R.H., Belle A., Bai O., Qi X., Ward K.R., Pfaffenberger M.P., Najarian K. A hierarchical method for removal of baseline drift from biomedical signals: Application in ECG analysis. Sci. World J. 2013
- [18] Monge-Álvarez J. Higuchi, Katz Fractal Dimension Measures, (2015) Poveznica: <https://www.mathworks.com/matlabcentral/fileexchange/50290-higuchi-and-katz-fractal-dimension-measures/content/Fractaldimensionmeasures/HiguchiFD.m>; pristupljeno: 25. siječnja 2024
- [19] M. Rahman and B. I. Morshed, "Extraction of Respiration Rate from Wrist ECG Signals," 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, (2021), Poveznica: <https://ieeexplore.ieee.org/document/9666489>; pristupljeno: 25. siječnja 2024
- [20] Šnajder J., Dalbello Bašić B., Strojno učenje, Fakultet elektrotehnike i računarstva, (2014) Poveznica: [https://www.fer.unizg.hr/_download/repository/StrojnoUcenje\[2\].pdf](https://www.fer.unizg.hr/_download/repository/StrojnoUcenje[2].pdf); pristupljeno 4. veljače 2024
- [21] Šnajder J., Dalbello Bašić B., Đurilić T., Merćep A., Strojno učenje Bilješke s predavanja, Fakultet elektrotehnike i računarstva, (2015/2016) Poveznica: https://www.fer.unizg.hr/_download/repository/SU-2015-Vrednovanje_modela.pdf; pristupljeno 4. veljače 2024
- [22] Begušić S., Dalbello Bašić B., Kostanjčar Z., Pintar D., Analiza glavnih komponenti, Multivarijatna analiza podataka, Fakultet elektrotehnike i računarstva, (2022/2023) Poveznica: https://www.fer.unizg.hr/_download/repository/MAP_5_Analiza_glavnih_komponenti.pdf; pristupljeno: 2. veljače 2024

Sažetak

Naslov: Određivanje krvnog tlaka iz elektrokardiograma metodama strojnog učenja

Sažetak: Ovim radom istražuje se upotreba strojnog učenja za određivanje krvnog tlaka iz elektrokardiogramskih (EKG) signala. Uključuje prikupljanje EKG podataka i krvnog tlaka, obradu signala, odabir značajki i klasifikaciju tlaka. Implementira metode su stabla odlučivanja, slučajne šume, potporni vektorski strojevi (SVM) i logistička regresija. Koje su korištene za klasifikaciju u 2 različita slučaja. U 2 klase: sniženi i povišeni krvni tlak. Te 3 klase: sniženi, normalni i povišeni krvni tlak. Rezultati ukazuju na mogućnosti ovakvih metoda ali nisu dovoljno precizni za široku upotrebu. Idući koraci su širenje baze podataka i implementacija naprednih metoda poput dubokog učenja.

Ključne riječi: krvni tlak, elektrokardiogram, strojno učenje, baza podataka, digitalna pred obrada, neinvazivno mjerenje, verifikacija, validacija, metoda, kategorizacija, algoritam, elektrokardiografski signal, filtriranje, segmentiranje, korelacija, sučelje, implementacija, procjena, klasifikacija, model, preciznost, točnost, analiza podataka, znanost o podacima, zdravstvo, medicinska dijagnostika, tehnologija.

Summary

Title: Determining Blood Pressure from Electrocardiograms Using Machine Learning Methods

Summary: This thesis investigates the use of machine learning for determining blood pressure from electrocardiographic (EKG) signals. It includes collecting EKG data and blood pressure, signal processing, feature selection, and pressure classification. Methods implemented are decision trees, random forests, support vector machines (SVM), and logistic regression. These were used for classification in 2 different cases. In 2 classes: lowered and elevated blood pressure. And 3 classes: lowered, normal, and elevated blood pressure. The results indicate the potential of such methods but are not accurate enough for widespread use. The next steps include expanding the database and implementing advanced methods such as deep learning.

Key words: blood pressure, electrocardiogram, machine learning, database, digital preprocessing, non-invasive measurement, verification, validation, method, categorization, algorithm, electrocardiographic signal, filtering, segmenting, correlation, interface, implementation, estimation, classification, model, precision, accuracy, data analysis, data science, healthcare, medical diagnostics, technology.