Estimacija krvnog pritiska koristeći fotopletizmografiju

# Uvod

U vreme kada sve više ljudi pati od hroničnog stresa i hipertenzije, praćenje i regulisanje krvnog pritiska postaje sve važnije. Iako veoma bitna metrika, krvni prisitak se često meri na invazivan način koji nije praktičan i uzima dosta vremena. Kao rešenje rešenje ovog problema napravljeni su mali i prenosivi uredjaji zasnovani na fotopletizmografiji (photoplethysmography tj. PPG). Medjutim, fotopletizmografija zahteva korišćenje statističkih ili metoda mašinskog učenja da bi se estimirao krvni pritisak. U skorije vreme, razvijeni su mnogli ML modeli ali bez konsenzusa oko metodologije i etalona. Da bi ovo promenili Gonzales et. al [1] su predstavili korektno particionisane setove podataka kao i rezultate predikcije sa odredjenim metrikama, koji služe kao etalon za ostale modele. Naš zadatak je bio da pokušamo da repliciramo rad i da pokušamo da istreniramo modele koje će nadmašiti postavljene standarde.

# Razrada

Visoki krvni pritisak ili hipertenzija povećava rizik od srčanog i moždanog udara, bolesti bubrega i oštećenja oka itd. Od njega boluje svaka druga osoba u Srbiji, a dijagnozu i propisanu terapiju dobije samo 33,9 odsto, podaci su Udruženja za hipertenziju Srbije [2]. Povišen krvni pritisak nazivaju i „tihim" ubicom jer nema neke posebne simptome[3]. Stoga je od velike važnosti da se hipertenzija otrkije na vreme. Klasični merači pritska nisu rešenje zato što nije realno za očekivati da neko konstantno navlači uredjaj i pumpa vazduh. Sa druge strane uredjaji koji koriste fotopletizmografiju nam omogućavaju da kontinualno i komotno merimo pritisak. Oni nisu toliko precizni, ali algoritmi mašinskog učenja imaju potencijal da ih unaprede.



Slika 1. PPG- uredjaj. Koristi se za merenje krvnog pritiska ali mu je primarna funkcija merenje kiseonika u krvi i njenog protoka

Postoje nekoliko načina da se estimira krvni pritisak od PPG mera:

PPT (Pulse Transit Time) bazirane metode koje koriste dva PPG uredjaja. Ove metode zahtevaju zasebnu kalibraciju za svakog pacijenta

Pristupi sa jednim PPG uredjajem koji ne zahtevaju kalibraciju Kombinovanje PPG i ARP (Arterial Blood Pressure) mera.

Mi ćemo se koncetristai na pristup sa jednim PPG uredjajem i to specifično na metode koje rade ekstrakciju atributa iz PPG talasa a zatim ih koriste za predikciju diskretne vrednosti krvnog pritiska. Razlog za to jeste što se ovaj metod pokazao najprezicniji na najvećem broju problema. Oni takodje ne zahtevaju veliki broj podataka da bi postigli dobre rezultate. Za ovu metodu, autori rada su omogućili i mere bitnosti atributa koje su dobili korišćenjem algoritama stabla odlučivanja. Mere bitnosti, kao i etaloni su odradjeni posebno za izračunavanje donjeg i gornjeg pritiska te ćemo i mi razvijate zasebne algoritme.

Biranje seta atributa je jedna od ključnih odluka u kreiranju modela i to iz dva razloga. Prvi razlog je da set atributa koje odaberemo utiče na kvalitet naših predikcija. Možemo da izaberemo premalo atributa i tako uskratimo modelu važne informacije a takodje možemo da uzmemo previše atributa, koji su često redudantni i beskorisni, i tako naštetimo modelu. Drugi razlog jeste to što će broj atributa uticati na memoriju i brzinu izvršavanja predikcija, što je veoma bitno kada imamo male uredjaje poput PPG.

Atributi koji su korišteni za predikovanje krvnog pritiska su:

point/time bazirani

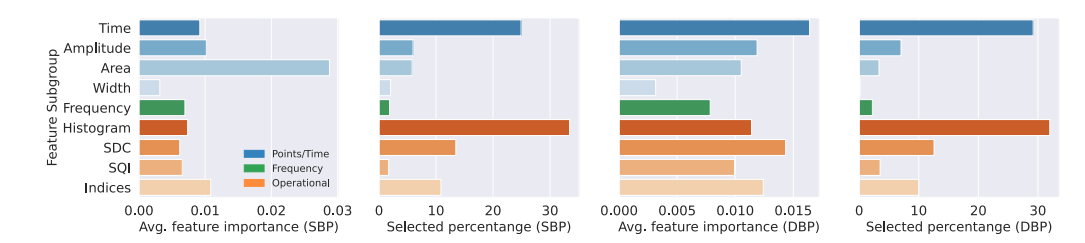
merenje proteklog vremena, amplitude, površine i širine izmedju tačaka od interesa srčanog kruga

Bazirani na fekfenciji Statistički atributi

Histogram

Slope Deviation Curve (SDC) Signal Quality Index (SQI)

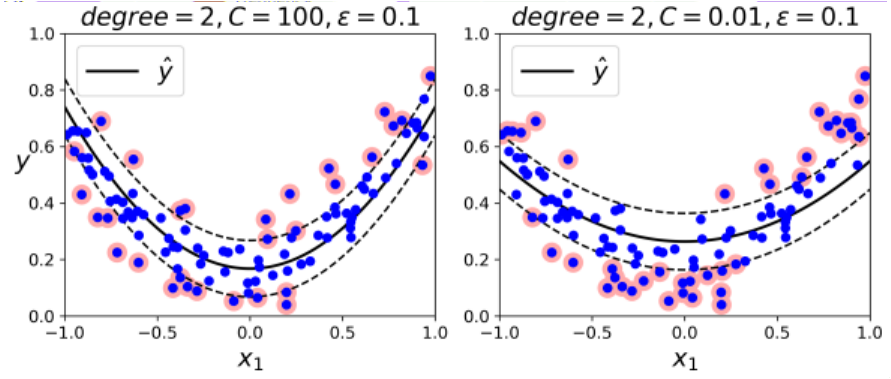
Nakon utvrdjivanja bitnosti atributa, tri najbitnija su ispali: proteklo vreme od vrha sistolnog pritiska do dihrotičkog udubljenja i dijastolnog porasta, i amplituda prvog izvoda. Autori rada su takodje pokazali relevantnost svake podgrupe. Medju atributima selektovanih od strane njihovih algoritama, grupe atributa sa najvecim procentom selekciju su histogram i vremenski bazirane grupe. Ipako mora se uzeti u obzir da ovaj rezultat može biti pristrasan zato što su grupe atributa različite veličine. Kada se uzme srednja bitnost atributa vremenski bazirani atributi ostaju medju najbitnijima dok histogram bazirani gube na vrednosti. SDC takodje ima pojedine korisne atribute dok su oni bazirani na frekfenciji i širini najmanje korisni.



Slika 2 Bitnost atributa. Prva i treća slika označavaju prosečnu bitnost atributa dok druga i četvrta prikazuju procenat selektovanih atributa

Gonzales je stavio poseban akcenat na particionisanje podataka zato što dosta objaveljnih radova ne odradi ovo kako treba i stoga dobijaju optimistične rezultate. Za razliku od klasične metode deljenja podataka, gde svaku torku nasumično delimo jednoj od tri grupe (train/dev/test), u slučaju kada radimo sa pacijentima moramo da vodimo računa da sve torke vezane za jednog pacijenta budu u istoj grupi. Ukoliko bi podaci za jednog pacijenta bili i u traning i test setu, to bi dalo veliku prednost algoritmu koju neće imati u realnom slučaju.

Kako bi probali da nadmašimo rezultate dobijene u prethodnom radu koristićemo Support Vector Regressor, Random Forest i Multi Layer Perceptron. Support Vector Regressor (SVR) je algoritam mašinskog učenja koji koristi potporne vektore kako bi pronašao što bolju granicu odlučivanja. Za razliku od Support Vector Classifier algoritma gde algoritam teži da fituje što širi "put", SVR pokušava da "put" bude što uži a da sve instance budu unutar njega. Random Forest se sastoji od ansambla stabala odlučivanja. Svako stablo radi sa nasumičnim setom atributa i daje izlaz koji se zatim usrednjava zajedno sa izlazima iz ostalih stabala i daje konačno rešenje. Multi Layer Perceptron je algoritam baziran na neuronskim mrežama koje koriste niz matričnih množenja i nelinearnih funkcija kako bi procesirali podatke.



Slika 3 Rezultat fitovanja dva SVR-a sa različitim parametrima. Cilj je da "put" tj. siprekidane linije budu što uže a da zahvate što više instanci.

# Rezultati

Treniranje smo odradili upotrebom kros validacije sa 5 predodredjenih particija. Koristili smo metodu GridSearchCV da bismo pronašli najbolje parametre. Parametre koje smo uzeli za "uži izbor" su oni koji se najčešće modifikuju. Takodje smo pokušali da nadjemo optimalan broj atributa koje svaki model treba da koristi. Rezultate koje prijavljujemo su dobijeni uzimanjem srednje vrednosti greške na test particijama a sama analiza je odradjena na tri dataset-a: PPGBP, Sensor, BCG.

## Sensor dataset

Na sensor datasetu najbolje se pokazao SVR model sa tim da i ostali modeli dali približno ista rešenja. Broj atributa koje su najbolji modeli koristili je u srednjem opsegu.

| **Model** | **Parametri** | **Broj Atributa** | **MAE** | **ME** | **SD** | **MASE** |
| --- | --- | --- | --- | --- | --- | --- |
| MLP | {'alpha': 0.01, 'hidden\_layer\_sizes': (50, 25)} | 50 | 7.73 | 0.63 | 9.45 | 93.65 |
| RF | {'max\_depth': 10, 'min\_samples\_split': 2, 'n\_estimators': 100} | 100 | 7.65 | -0.06 | 7.29 | 92.60 |
| SVR | {'C': 100, 'epsilon': 0.3, 'kernel': 'rbf'} | 20 | 7.56 | -1.73 | 9.65 | 91.61 |

Tabela 1 Rezultati predikcije dijastolnog pritiska na Sensor dataset-u

| **Model** | **Parametri** | **Broj Atributa** | **MAE** | **ME** | **SD** | **MASE** |
| --- | --- | --- | --- | --- | --- | --- |
| MLP | {'alpha': 0.01, 'hidden\_layer\_sizes': (25, 25)} | 50 | 15.98 | -0.11 | 19.64 | 90.91 |
| RF | {'max\_depth': None, 'min\_samples\_split': 10, 'n\_estimators': 100} | 100 | 15.65 | 0.06 | 8.99 | 89.02 |
| SVR | {'C': 100, 'epsilon': 0.3, 'kernel': 'linear'} | 20 | 16.09 | -0.49 | 20.07 | 91.58 |

Tabela 2 Rezultati predikcije sistolnog pritiska na Sensor dataset-u

| **Model** | **MAE\_SP** | **ME\_SP** | **SE\_SP** | **MASE\_SP** | **MAE\_DP** | **ME\_DP** | **SE\_DP** | **MASE\_DP** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MLP | **15.98** | -0.11 | 19.64 | 90.91 | **7.73** | 0.63 | 9.45 | 93.65 |
| MLP Iz rada | 16.03 | -0.50 | 20.10 | 91.03 | 7.77 | -0.19 | 10.04 | 94.05 |
| RF | **15.65** | 0.06 | 8.99 | 89.02 | **7.65** | -0.06 | 7.29 | 92.60 |
| RF Iz rada | 15.86 | -0.12 | 19.85 | 90.08 | 7.66 | -0.03 | 9.86 | 92.63 |
| SVR | 16.09 | -0.49 | 20.07 | 91.58 | 7.56 | -1.73 | 9.65 | 91.61 |
| SVR Iz rada | **15.60** | -0.00 | 19.68 | 88.62 | **7.50** | -1.45 | 9.81 | 90.76 |

Tabela 3 Poredjenje naših sa rezultatima iz rada na Sensor dataset-u

## BCG dataset

Kod BCG seta SVR se opet pokazao kao najbolji model ali je ovaj put dosta precizniji u odnosu na RF i MLP. Na ovom setu naš SVR model je dao najtačniju analizu kada je u pitanju dijastolni pritisak.

| **Model** | **Parametri** | **Broj Atributa** | **MAE** | **ME** | **SD** | **MASE** |
| --- | --- | --- | --- | --- | --- | --- |
| MLP | {'alpha': 0.1, 'hidden\_layer\_sizes': (50,)} | 50 | 8.01 | -0.20 | 5.35 | 108.00 |
| RF | {'max\_depth': 10, 'min\_samples\_split': 2, 'n\_estimators': 100} | 100 | 7.75 | -0.02 | 2.58 | 104.58 |
| SVR | {'C': 1, 'epsilon': 0.3, 'kernel': 'linear'} | 20 | 6.89 | -0.20 | 6.92 | 92.99 |

Tabela 4 Rezultati predikcije dijastolnog pritiska na BCG dataset-u

| **Model** | **Parametri** | **Broj Atributa** | **MAE** | **ME** | **SD** | **MASE** |
| --- | --- | --- | --- | --- | --- | --- |
| MLP | {'alpha': 0.1, 'hidden\_layer\_sizes': (50, 25)} | 50 | 13.14 | -0.97 | 7.95 | 118.16 |
| RF | {'max\_depth': 10, 'min\_samples\_split': 5, 'n\_estimators': 50} | 100 | 12.23 | 0.04 | 4.03 | 109.97 |
| SVR | {'C': 0.1, 'epsilon': 0.1, 'kernel': 'rbf'} | 20 | 11.52 | -0.40 | 13.99 | 103.56 |

Tabela 5 Rezultati predikcije sistolnog pritiska na BCG dataset-u

| **Model** | **MAE\_SP** | **ME\_SP** | **SE\_SP** | **MASE\_SP** | **MAE\_DP** | **ME\_DP** | **SE\_DP** | **MASE\_DP** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MLP | 13.14 | -0.97 | 7.95 | 118.16 | 8.01 | -0.20 | 5.35 | 108.00 |
| MLP Iz rada | **12.98** | -0.27 | 16.35 | 105.50 | **7.14** | 0.03 | 9.28 | 90.24 |
| RF | **12.23** | 0.04 | 4.03 | 109.97 | **7.75** | -0.02 | 2.58 | 104.58 |
| RF Iz rada | 12.88 | -1.46 | 17.75 | 104.72 | 7.89 | -0.01 | 10.44 | 99.77 |
| SVR | 11.52 | -0.40 | 13.99 | 103.56 | **6.89** | -0.20 | 6.92 | 92.99 |
| SVR Iz rada | **11.45** | -0.79 | 15.56 | 93.07 | 7.34 | 0.01 | 9.88 | 92.75 |

Tabela 6 Poredjenje naših sa rezultatima iz rada na BCG dataset-u

## PPGBP dataset

Modeli dobijeni na PPGBP setu se razlikuju od ostalih zato što oni najbolje rade sa malim broj atributa. SVR je imao najbolje rezultate sa svega četiri atributa dok je Random Forest, koji je daje najbolje rezultate od svih modela na ovom setu podataka, ima samo 12 feature-a.

| **Model** | **Parametri** | **Broj Atributa** | **MAE** | **ME** | **SD** | **MASE** |
| --- | --- | --- | --- | --- | --- | --- |
| MLP | {'alpha': 0.1, 'hidden\_layer\_sizes': (50, 25)} | 8 | 8.98 | -0.55 | 11.00 | 101.59 |
| RF | {'max\_depth': None, 'min\_samples\_split': 10, 'n\_estimators': 100} | 12 | 7.99 | -0.01 | 5.11 | 90.36 |
| SVR | {'C': 100, 'epsilon': 0.1, 'kernel': 'linear'} | 4 | 8.08 | -1.29 | 10.17 | 91.38 |

Tabela 7 Rezultati predikcije dijastolnog pritiska na PPGBP dataset-u

| **Model** | **Parametri** | **Broj Atributa** | **MAE** | **ME** | **SD** | **MASE** |
| --- | --- | --- | --- | --- | --- | --- |
| MLP | {'alpha': 0.01, 'hidden\_layer\_sizes': (50, 25)} | 8 | 17.17 | 1.01 | 19.60 | 104.95 |
| RF | {'max\_depth': 20, 'min\_samples\_split': 10, 'n\_estimators': 100} | 12 | 12.89 | -0.06 | 8.68 | 78.79 |
| SVR | {'C': 100, 'epsilon': 0.3, 'kernel': 'linear'} | 4 | 14.32 | -1.29 | 18.20 | 87.50 |

Tabela 8 Rezultati predikcije sistolnog pritiska na PPGBP dataset-u

| **Model** | **MAE\_SP** | **ME\_SP** | **SE\_SP** | **MASE\_SP** | **MAE\_DP** | **ME\_DP** | **SE\_DP** | **MASE\_DP** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MLP | 17.17 | 1.01 | 19.60 | 104.95 | 8.98 | -0.55 | 11.00 | 101.59 |
| MLP Iz rada | **13.38** | -0.13 | 17.09 | 81.69 | **8.21** | -0.16 | 10.40 | 92.77 |
| RF | **12.89** | -0.06 | 8.68 | 78.79 | **7.99** | -0.01 | 5.11 | 90.36 |
| RF Iz rada | 13.17 | 0.02 | 16.81 | 80.42 | 8.12 | 0.19 | 10.17 | 91.76 |
| SVR | 14.32 | -1.29 | 18.20 | 87.50 | 8.08 | -1.29 | 10.17 | 91.38 |
| SVR Iz rada | **13.15** | -0.64 | 17.05 | 80.29 | **8.04** | -0.22 | 10.14 | 90.90 |

Tabela 9 Poredjenje naših sa rezultatima iz rada na PPGBP dataset-u

# Zaključak

Kroz istraživanje ovog rada, dublje sam shvatio problematiku koju hipertenzija prouzrokuje kao i važnost u pravovremenom otrkiću i prevenciji. U rešavanju ovih problema, fotopletizmografija ima veliki potencijal, medjutim njen uspeh zavisi u mnogome od kvaliteta ML modela. Gonzales et. al. su dali ogroman doprinos na ovom polju, tako što su postavili etalon za poredjenje budućih modela. Rešenja koja smo mi razvili, su približna, a u nekim slučajevima i bolja od ustanovljenih rezultata.

Nadam se da će ovaj rad poslužiti za dalje istraživanje i razmatranje ove važne problematike.

# Reference

1. González, S., Hsieh, WT. & Chen, T.PC. A benchmark for machine-learning based non-invasive blood pressure estimation using photoplethysmogram. *Sci Data* **10**, 149 (2023). <https://doi.org/10.1038/s41597-023-02020-6>
2. BBC News na sprskom, Svetski dan borbe protiv visokog krvnog pritiska: Šta je [hipertenzija i zašto se naziva „tihim ubicom" https://www.bbc.com/serbian/lat/svet- 61477116](https://www.bbc.com/serbian/lat/svet-61477116)
3. Udruženje za hipertenziju Srbije <https://www.uhsrb.rs/hipertenzija-u-srbiji/>