

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 297

**DETEKCIJA SRČANIH ŠUMOVA IZ SNIMKI
FONOKARDIOGRAMA METODAMA DUBOKOG UČENJA**

Marin Matjanec

Zagreb, veljača 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 297

**DETEKCIJA SRČANIH ŠUMOVA IZ SNIMKI
FONOKARDIOGRAMA METODAMA DUBOKOG UČENJA**

Marin Matjanec

Zagreb, veljača 2024.

DIPLOMSKI ZADATAK br. 297

Pristupnik: **Marin Matjanec (0036514463)**

Studij: Računarstvo

Profil: Računarska znanost

Mentor: izv. prof. dr. sc. Alan Jović

Zadatak: **Detekcija srčanih šumova iz snimki fonokardiograma metodama dubokog učenja**

Opis zadatka:

Otkucaji srca generiraju zvukove koji se auskultacijom pomoću digitalnog stetoskopa mogu snimati i spremati u signal fonokardiograma (engl. phonocardiogram, PCG). Signali fonokardiograma daju vrijedne informacije koje se odnose na srčane zaliske. Putem PCG-a moguće je otkriti srčane šumove koji sugeriraju postojanje srčanih bolesti. Model strojnog učenja koji na temelju snimke fonokardiograma detektira određene srčane šumove može pomoći liječnicima kod dijagnosticiranja srčanih oboljenja. U ovom diplomskom radu potrebno je razviti programski sustav s implementacijom modela strojnog učenja koji će analizom signala PCG-a detektirati prisutnost srčanih šumova. Sustavom je potrebno podržati izlučivanje značajki PCG-a koje se potom koriste u modelima strojnog i dubokog učenja. U radu je potrebno implementirati nekoliko postupaka strojnog i dubokog učenja s ciljem usporedbe uspješnosti modela. Snimke PCG-a treba preuzeti iz slobodno dostupnog skupa podataka CirCor DigiScope (<https://physionet.org/content/circor-heart-sound/1.0.0/>), koji sadržava velik broj snimki PCG-a većinom pedijatarske populacije. U diplomskom radu također je potrebno usporediti izgrađene modele s rezultatima srodnih metoda drugih istraživača koji su objavili rezultate na istom problemu.

Rok za predaju rada: 9. veljače 2024.

Sadržaj

| | |
|---|----|
| Uvod | 3 |
| 1. Fonokardiografija i srčani šumovi | 4 |
| 1.1. Kardiovaskularne bolesti | 4 |
| 1.2. Auskultacija srca..... | 5 |
| 1.3. Zvuk srca | 6 |
| 2. Skup podataka CirCor DigiScope..... | 12 |
| 2.1. Važnost dobrog skupa podataka u fonokardiografiji..... | 12 |
| 2.2. Kampanja i prikupljanje podataka | 13 |
| 2.3. Demografski i klinički podaci | 14 |
| 2.4. Zvukovi srca i njihovo označavanje | 14 |
| 3. Metode strojnog učenja | 17 |
| 3.1. Izlučivanje značajki iz audio zapisa | 17 |
| 3.1.1. Fourierova transformacija..... | 18 |
| 3.1.2. Kepstralni koeficijenti mel-frekvencije | 21 |
| 3.2. Logistička regresija..... | 23 |
| 3.3. Naivni Bayesov klasifikator | 26 |
| 3.4. Ansambl slučajna šuma | 27 |
| 3.5. XGBoost | 29 |
| 4. Konvolucijski i povratni modeli | 32 |
| 4.1. Duboko učenje | 32 |
| 4.2. Konvolucijski modeli | 33 |
| 4.2.1. Jednodimenzionalni konvolucijski model | 34 |
| 4.2.2. Dvodimenzionalni konvolucijski model..... | 36 |
| 4.2.3. AlexNet..... | 38 |
| 4.2.4. ResNet | 40 |

| | | |
|--------|--|----|
| 4.2.5. | Metoda podjele PCG-a na srčane cikluse | 42 |
| 4.3. | Povratni modeli..... | 46 |
| 4.3.1. | Ćelija s dugom kratkoročnom memorijom..... | 47 |
| 4.3.2. | Propusna povratna ćelija..... | 48 |
| 4.4. | Kombinacija konvolucijskih i povratnih slojeva | 49 |
| 5. | Transformerski modeli | 51 |
| 5.1. | Wav2vec 2.0 | 52 |
| 5.2. | HuBERT | 53 |
| 5.3. | Audio Spectrogram Transformer | 54 |
| 5.4. | Whisper..... | 55 |
| 6. | Rasprava i rezultati | 58 |
| | Zaključak | 62 |
| | Literatura | 64 |
| | Sažetak..... | 68 |
| | Summary..... | 69 |
| | Skraćenice..... | 70 |

Uvod

Kardiovaskularne se bolesti svake godine nalaze na vrhu ljestvice najsmrtonosnijih bolesti na svijetu. Procjenjuje se da urođene kardiovaskularne bolesti pogađaju oko 1% novorođenčadi. Takve bolesti bitno je što prije otkriti, jer mogu utjecati na druga teška oboljenja. Nedostatak rane dijagnoze urođenih bolesti srca predstavlja veliki problem u siromašnim zemljama s visokim stopama nataliteta gdje često nema dovoljno liječnika. Neinvazivna tehnika procjene mehaničkog djelovanja srca može dati pravovremene informacije o prirođenim i stečenim bolestima srca u djece, koje se često karakteriziraju neuobičajenim zvukom, srčanim šumom (Reyna et al., 2022).

Kroz praksu se pokazalo kako je auskultacija jedna od najisplativijih tehnika za otkrivanje i prepoznavanje urođenih ili stečenih srčanih oboljenja (Oliveira et al., 2022.). Auskultacija stetoskopom se provodi vrlo jednostavno i neinvazivno, a fokusira se na detekciji abnormalnih šumova u zvuku srca. Danas postoje digitalni stetoskopi koji zvuk srca snimaju u zapise fonokardiograma (engl. *phonocardiogram*, PCG).

Budući da mnoge zemlje svijeta imaju poteškoća u dijagnosticiranju i liječenju urođenih i stečenih srčanih bolesti, računalnom analizom i razvojem algoritama umjetne inteligencije može se pomoći liječnicima u donošenju bržih i kvalitetnijih dijagnoza.

Cilj ovog diplomskog rada je razviti algoritam umjetne inteligencije za detekciju prisutnosti srčanih šumova u PCG-u, koji će moći biti koristan u medicini za pomoć pri dijagnosticiranju srčanih oboljenja. Ideja je ispitati i usporediti više različitih principa strojnog i dubokog učenja, s ciljem razvoja što uspješnijeg modela.

U nastavku su opisani fonokardiografija i srčani šumovi, jer je njihovo razumijevanje bitan dio rješavanja problema detekcije šuma i razvoja inteligentnih modela. Dane su informacije o korištenom skupu podataka CirCor Digiscope. Prikazani su i detaljno opisani modeli strojnog i dubokog učenja korišteni u ovom istraživanju. Rezultati su izneseni na kraju rada zajedno s usporedbom rezultata drugih radova sa srodnim temama.

1. Fonokardiografija i srčani šumovi

U ovom poglavlju dan je kratki opis kardiovaskularnih bolesti, fonokardiograma, zvukova koje srce proizvodi u svom ciklusu i šumova koji, ako postoje, mogu ukazivati na neke od kardiovaskularnih bolesti. Obrazložit će se važnost i iskoristivost fonokardiograma i auskultacije srca kod ranog otkrivanja srčanih oboljenja.

1.1. Kardiovaskularne bolesti

Svjetska zdravstvena organizacija redovito objavljuje statističke podatke o zdravlju ljudi diljem svijeta. Prema njihovim najnovijim podacima iz 2023., kardiovaskularne bolesti najveći su uzročnik smrti u svijetu. Kardiovaskularne bolesti svake godine oduzmu oko 17,9 milijuna života, što čini čak 31% od svih smrtnih uzročnika (Mattingly, 2024). Tih 17,9 milijuna je više od broja umrlih od malignih bolesti, respiratornih bolesti i dijabetesa zajedno, koje čine tri sljedeće skupine najsmrtonosnijih nezaraznih bolesti (WHO, 2023).

Kardiovaskularna bolest je zajednički pojam koji se koristi za definiranje heterogene skupine poremećaja srca i krvnih žila, kao što su bolest koronarnih arterija (engl. *coronary artery disease*, CAD), bolest srčanih zalistaka (engl. *valvular heart disease*, VHD) ili prirođene srčane mane (engl. *congenital heart disease*, CHD) (Libby et al., 2007; Oliveira et al., 2022).

Osim što utječu na smrtnost, kardiovaskularne bolesti smanjuju kvalitetu života, ozbiljno povećavaju morbiditet i uzrokuju cjeloživotne invalidnosti i snažno povećavaju učestalost bolničkih prijema, u konačnici povećavajući ekonomski teret kardiovaskularnih bolesti u zdravstvenim sustavima (Gheorghe et al., 2018; Oliveira et al., 2022). Većina stanovništva u nerazvijenim zemljama i zemljama u razvoju ima ograničen pristup primarnom zdravstvenom osiguranju, zbog čega dijagnoza i liječenje kardiovaskularnih bolesti često kasne, što može dovesti do rane smrti oboljelih. Nadalje, kardiovaskularne bolesti doprinose osiromašenju takvih zemalja zbog zdravstvenih troškova. To nameće dodatni teret na gospodarstva zemalja s niskim do srednjim dohotkom (Murphy et al., 2020; Oliveira et al., 2022). Iako je broj slučajeva kardiovaskularnih bolesti u SAD-u i Europi visok, procjenjuje se da se više od 75 % smrtnih slučajeva povezanih s kardiovaskularnim bolestima događa u zemljama s niskim i srednjim dohotkom, što ukazuje na snažnu povezanost sa socioekonomskim statusom (Lea et al., 2006; Oliveira et al., 2022).

U razvijenim zemljama češća je bolest koronarnih arterija, dok kod nerazvijenih zemalja i zemalja u razvoju češće su bolesti srčanih zalistaka i prirodene srčane mane, prvenstveno zbog nedostatka pregleda u trudnoći i pristupa zdravstvenoj skrbi (Soler-Soler & Galve, 2000; Oliveira et al., 2022).

Ovi podaci ukazuju na važnost ranog i preciznog otkrivanja i adekvatnog liječenja bolesti srčanih zalistaka i prirodnih srčanih bolesti koje se najčešće javljaju u slabije razvijenim zemljama i zemljama u razvoju. Pregled stetoskopom kojim se zvuk srca zapisuje u oblik fonokardiograma može biti značajan za doprinos olakšavanju otkrivanja ovakvih bolesti.

1.2. Auskultacija srca

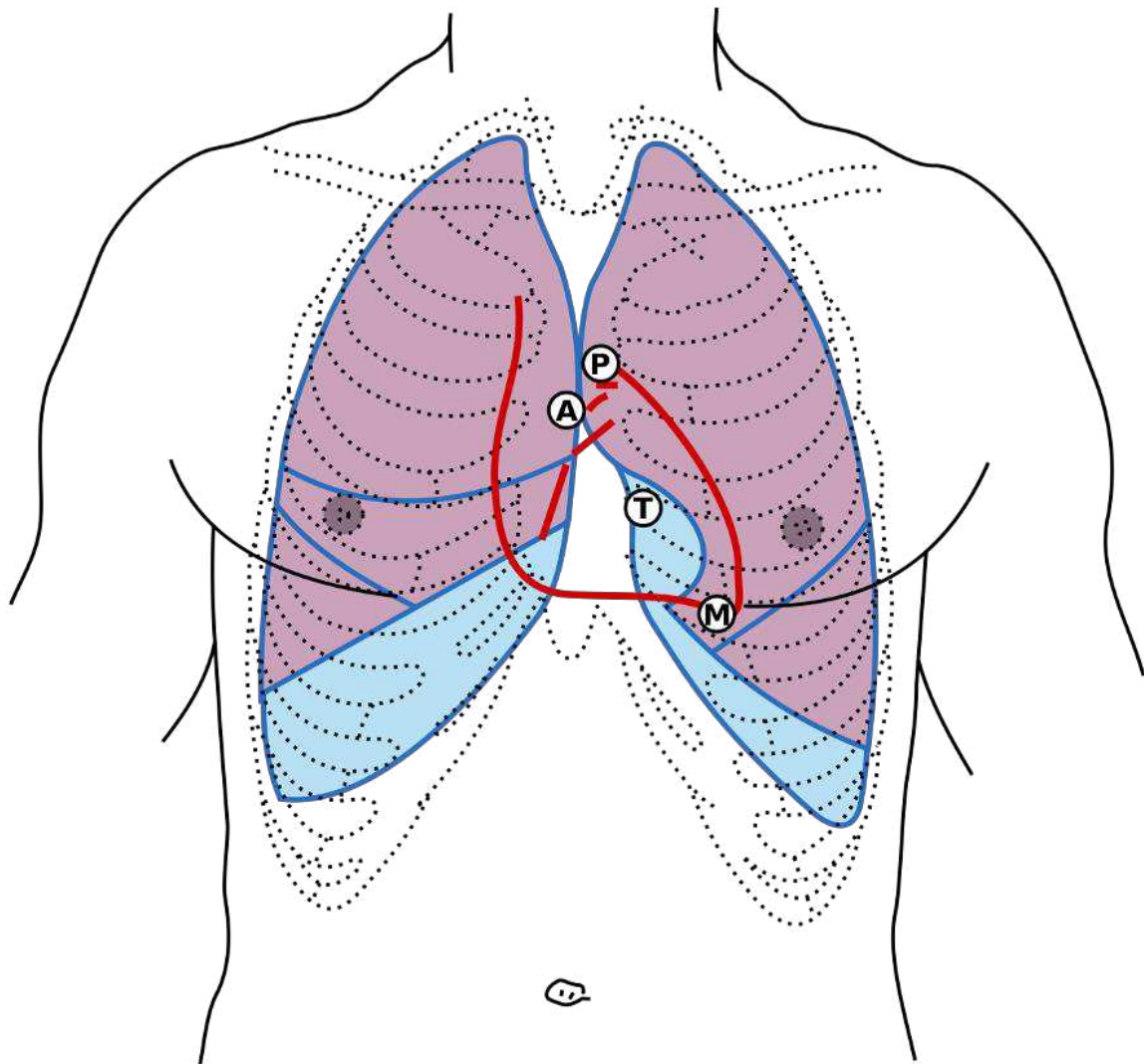
Iako se često koriste višestruki načini ispitivanja srca, uključujući elektrokardiografiju i ehokardiografiju uz napredne tehnike snimanja uključujući tomografiju i magnetsku rezonanciju, još jedan alat koji može pomoći u dijagnozi prirodnih srčanih bolesti je fonokardiografija (Burns et al., 2022).

Unatoč razvoju naprednih tehnika za praćenje rada srca, auskultacija ostaje važan prvi izbor kao jedan neinvazivan, jednostavan i troškovno učinkovit alat. Auskultacija daje informacije o mehaničkoj aktivnosti srca i stanju srčanih zalistaka. Takve aktivnosti generiraju srčane zvukove koji se snimaju i spremaju u audio-signal PCG-a (Oliveira et al., 2022).

Trenutna literatura sugerira da je PCG učinkovit, osjetljiv i specifičan u dijagnosticiranju strukturalne bolesti srca kod pedijatarske populacije. Iako podaci odražavaju promjenjivi uspjeh u identificiranju specifičnih lezija, PCG je općenito vrlo učinkovit u otkrivanju prirodnih srčanih bolesti i razlikovanju nevinih od patoloških šumova. Osim toga, podaci sugeriraju da PCG može biti korisniji prediktor urođenih srčanih bolesti od konvencionalne auskultacije (Burns et al., 2022).

Normalni srčani zvukovi uglavnom proizlaze iz vibracija koje proizvodi otvaranje i zatvaranje srčanih zalistaka u svakom srčanom ciklusu i kretanje krvi u arterijama. Idealno mjesto auskultacije određeno je anatomskim položajem srčanih zalistaka u odnosu na stijenku prsnog koša. Stoga je za učinkovitu auskultaciju potrebno postavljanje stetoskopa na mjesta označena na Slika 1.1 (Oliveira et al., 2022).

M označava mitralno polje, T trikuspidno polje, A aortno polje i P pulmonalno polje.



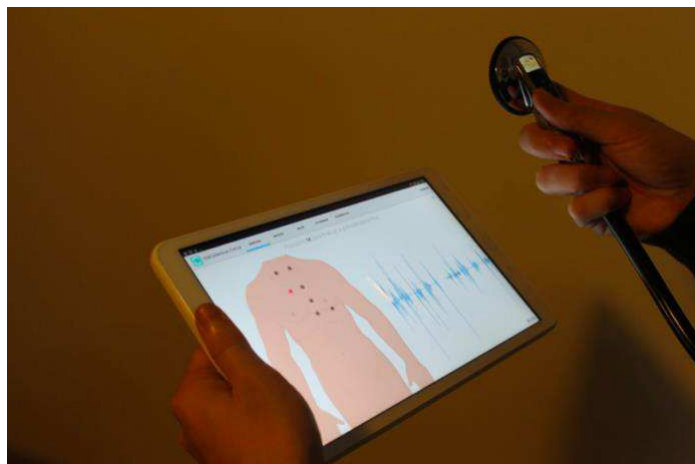
Slika 1.1 Auskultacijski položaji (modificirano od Gray, 1918)

Auskultacija srca je često korištena tehnika za detektiranje fizičkih anomalija kardiovaskularnog sustava, ali postavljanje dijagnoze na temelju srčanih zvukova korištenjem samo stetoskopa vještina je za čije svladavanje može trebati podosta vremena. Točnost i učinkovitost dijagnoze temeljene na auskultaciji srca i detekciji srčanih šumova može se značajno poboljšati korištenjem digitalne obrade signala i tehnika umjetne inteligencije za istraživanje PCG signala (Subasi, 2019).

1.3. Zvuk srca

Zvučni signal srca je tradicionalni i jedan od osnovnih biomedicinskih signala, jer je stetoskop primarni instrument koji liječnici koriste prilikom pregleda pacijenata. Kako se zvuk proizvodi i prenosi vibracijama, te vibracije su sačinjene od niza valova koji se dijele

na kompresije koje su područja povećanog tlaka i naknadna razrjeđivanja koja su područja sniženog tlaka. Takvi valovi prenose se kroz čvrste, tekuće ili plinovite medije. Općenito vrijedi da su lakoća i brzina prijenosa zvučnih valova obrnuto proporcionalne gustoći medija u kojem putuju. Liječnici obično koriste stetoskop za prijenos kardiovaskularnih zvukova od stijenke prsnog koša do ušiju. Za razliku od klasičnih stetoskopa, fonokardiografija pomoću digitalnih stetoskopa omogućuje otkrivanje zvukova i šumova nečujnih ljudskom uhu i stvaranje trajnih zapisa tih događaja (Ranganathan et al., 2007; Subasi, 2019). Slika 1.2 prikazuje jedan moderan digitalni stetoskop Littmann 3000 koji se koristi za snimanje zvuka srca i izradu fonokardiograma.



Slika 1.2 Littmann 3000 stetoskop (Oliveira et al., 2022)

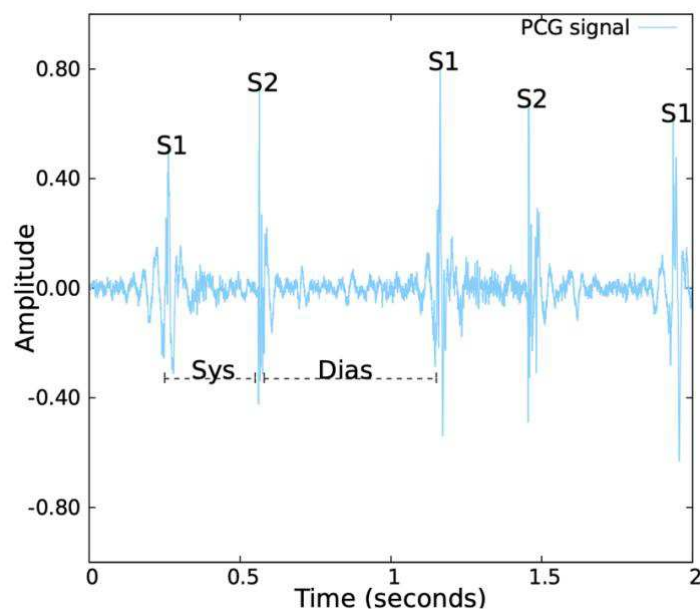
Svi srčani zvukovi nastaju naglim zaustavljanjem ili značajnim usporavanjem pokretne mase krvi. Jačina srčanog zvuka ovisit će o količini energije koju je sadržavala pokretna masa krvi. Naglo usporavanje krvi uzrokuje disipaciju energije, što proizvodi vibracije koje utječu na susjednu krvnu masu. Postoje razni čimbenici koji utječu na ubrzanje i usporavanje krvne mase koja je zaslužna za stvaranje nekoliko srčanih zvukova (Oweis et al., 2014; Subasi, 2019).

Dva najglasnija srčana zvuka su prvi (S1) i drugi (S2). Vremenski interval između S1 i S2 naziva se sistola, dok se vremenski interval između S2 i S1 naziva dijastola. Sistola je trenutak u srčanom ciklusu kad je srčani mišić u fazi kontrakcije i istiskanja krvi, a dijastola je trenutak opuštanja srčanog mišića i ulaska krvi u srce. Normalni srčani zvukovi koje proizvode srčani zalisci su niskofrekventni prolazni signali, dok su abnormalni srčani

zvukovi, kao što su srčani šumovi, visokofrekventni zvukovi nalik na buku. (Baguet et al., 2012; Subasi, 2019).

Prvi srčani zvuk (S1) nastaje vibracijama izazvanim zatvaranjem mitralnog i trikuspidnog zaliska na početku sistole. S1 se čuje na stijenci prsnog koša, a formiraju ga mitralna i trikuspidna komponenta. Iako mitralna komponenta S1 dolazi ranije i veće je glasnoće, dok je osoba u fiziološkom stanju mirovanja, obje se komponente javljaju dovoljno blisko, zbog čega je njihovo razlikovanje često otežano. Drugi srčani ton (S2) nastaje zatvaranjem aortnog i plućnog zaliska na početku dijastole. S2 također čine dvije komponente, od kojih je aortna komponenta glasnija i javlja se ranije od pulmonalne. To se događa zbog tlaka koji je u aorti veći nego u plućnoj arteriji. Za razliku od S1, u normalnim fiziološkim uvjetima može se razlikovati zvuk zatvaranja aortnog i plućnog zaliska. To razlikovanje je moguće, jer je povećanje tlaka u plućnoj arteriji, pa tako posljedično i zatvaranje plućnog zaliska odgođeno zbog povećanja venskog povrata tijekom udisaja (Dornbush & Turnquest, 2019; Oliveira et al., 2022).

Slika 1.3 prikazuje jedan normalan PCG signal na kojem su označeni srčani zvukovi S1 i S2, a također i periodi sistole i dijastole koji su određeni tim zvukovima.



Slika 1.3 Normalni PCG signal (Oliveira et al., 2022)

U nekim situacijama može se čuti treći srčani zvuk (S3), povezan s naglim prekidom faze brzog punjenja klijetke. Budući da su klijetke pune krvi i njihove stijenke su opušteno tijekom ovog dijela dijastole, vibracije S3 su vrlo niske frekvencije. U kasnoj dijastoli

povremeno se može čuti četvrti srčani zvuk (S4), nastao kontrakcijama pretklijetke koja prenosi krv u proširenu klijetku. Osim ovih zvukova rijetko se čuju klikovi i škljocanje srčanih zalistaka (Rangayyan, 2015).

Dodatni srčani zvukovi koji nastaju zbog turbulentnog protoka krvi i brzih pokreta srčanih struktura često su povezani sa srčanim šumovima (engl. *heart murmurs*). Takvi šumovi mogu biti povezani s fiziološkim ili patološkim stanjima, ovisno o njihovom vremenu u srčanom ciklusu, intenzitetu, promjeni intenziteta tijekom srčanog ciklusa, obliku signala, visini, mjestu auskultacije, zračenju, ritmu i odgovoru na manevre fizikalnog pregleda (Dornbush & Turnquest, 2019).

Srčani šumovi nastaju turbulencijama u protoku krvi kroz uske srčane zaliske. Urođene srčane mane ili otkrivene bolesti srčanih zalistaka općenito su izvor abnormalnih srčanih šumova. Stenoza aorte, aortna regurgitacija, regurgitacija mitralnog zaliska i mitralna stenoza su među najčešće viđenim patološkim vrstama šumova. Stenoza je nedovoljno otvaranje srčanog zaliska gdje dolazi do djelomičnog ograničenja protoka krvi između zalistaka, a regurgitacija je nedovoljno zatvaranje srčanog zaliska pri kojemu dolazi do djelomičnog propuštanja krvi unatrag kroz zaliske (Pollock & Makaryus, 2020; Oliveira et al., 2020).

U nastavku su navedeni opisi šumova uobičajenih bolesti zalistaka koji su opisani u (Pollock & Makaryus, 2020; Oliveira et al., 2022).

1. **Aortna stenoza:** Aortna stenoza je suženje aortnog zaliska i često se javlja kao posljedica kalcifikacije (akumulacije kalcijevih soli) zaliska. Tijekom auskultacije, stenoza aorte stvara oštar sistolički šum čija se glasnoća postepeno povećava, pa se postepeno smanjuje. Taj šum se najbolje čuje na gornjoj desnoj granici prsne šupljine, a također se širi i na karotidne arterije.
2. **Aortna regurgitacija:** Aortna regurgitacija je posljedica nedovoljnog zatvaranja aortnog zaliska, što uzrokuje i omogućuje da se krv na turbulentan način vrati natrag u desnu klijetku. Inače se pojavljuje zbog proširenja korijena aorte, bikuspidalnog aortnog zaliska i kalcificiranih listića. Tijekom auskultacije, aortnu regurgitaciju označava dijastolički šum koji je puhajućeg karaktera čija se glasnoća postepeno smanjuje. Takav šum najbolje se čuje na donjoj lijevoj granici prsne šupljine.
3. **Mitralna stenoza:** Mitralna stenoza je suženje mitralnog zaliska. Takvo suženje zaliska otežava prolazak krvi iz lijeve pretklijetke u lijevu klijetku. Tijekom

auskultacije, mitralna stenoza stvara dijastolički šum, koji se najbolje čuje na vrhu srca.

4. **Mitralna regurgitacija:** Mitralna regurgitacija omogućuje protok krvi natrag u lijevu pretkljetku, uglavnom zbog neispravnog zatvaranja ili oštećenja listića mitralnog zaliska. Kod auskultacije, mitralna regurgitacija rezultira sistoličkim šumom, koji se najbolje čuje na vrhu srca, ali se širi i u lijevu aksilu (regija ispod ramenog zgloba gdje se ruka spaja s ramenom).
5. **Prolaps mitralnog zaliska:** Prolaps mitralnog zaliska karakterizira izbočenje listića u lijevu pretkljetku, što onemogućuje ravnomjerno zatvaranje zaliska. U auskultaciji, prolaps mitralnog zaliska karakterizira rani sistolički klik koji se najbolje čuje na vrhu srca, a često nakon toga slijedi i šum u kasnijoj fazi sistole.
6. **Stenoza pulmonalnog zaliska:** Stenoza pulmonalnog zaliska, kako samo ime kaže, je suženje pulmonalnog zaliska. U auskultaciji opisuje ju sistolički šum čija se glasnoća postepeno povećava pa smanjuje. Najglasnije se čuje na gornjoj lijevoj granici prsne šupljine.
7. **Trikuspidna stenoza:** Trikuspidna stenoza je suženje u trikuspidnom zalisku. U auskultaciji, trikuspidna stenoza rezultira dijastoličkim šumom, koji se najbolje čuje na donjoj lijevoj granici prsne šupljine.
8. **Trikuspidna regurgitacija:** Trikuspidna regurgitacija obično je uzrokovana vegetativnim rastom ispod listića trikuspidnog zaliska, što uzrokuje njihovo propadanje, dopuštajući krvi da utječe natrag u desnu pretkljetku. U auskultaciji, trikuspidna regurgitacija opisuje se kao sistolički šum koji se najbolje čuje na donjoj lijevoj granici prsne kosti.
9. **Defekti septuma:** Defekti septuma su urođene prirode i definirani su kao oštećenja, puknuća ili prekidi u septumu pretkljetki ili u septumu kljetki. Takvi defekti omogućuju da se krv miješa i kreće između kljetki i pretkljetki. Defekt septuma pretkljetke u auskultaciji obično se očituje kao glasan i širok S1, fiksni podijeljeni S2 srčani ton, najglasniji na gornjoj lijevoj granici prsne kosti. Defekti septuma kljetke često stvaraju holosistolički šum, koji se najbolje čuje na vrhu srca. Manji defekti su glasniji i oštrije su kvalitete, dok su veliki tiši, ali uzrokuju jače simptome.
10. **Hipertrofična opstruktivna kardiomiopatija:** Hipertrofična opstruktivna kardiomiopatija je nasljedna bolest miokarda u kojoj dolazi do hipertrofičnih promjena kod srčanog mišića. Tijekom auskultacije, hipertrofična opstruktivna kardiomiopatija

predstavlja sistolički šum koji djeluje kao izbacivanje, a najbolje se čuje između vrha srca i lijeve granice prsne kosti.

Iako su najzastupljenije, ovo su samo neke od bolesti koje se jasno i precizno mogu dijagnosticirati analizom PCG signala, što ukazuje na važnost ovog signala kod dijagnoze nasljednih bolesti krvožilnog sustava i bolesti srčanih zalistaka koje pogađaju velik broj svjetskog stanovništva.

2. Skup podataka CirCor DigiScope

Podaci za ovo istraživanje korišteni su iz slobodno dostupnog skupa podataka CirCor DigiScope Dataset. Ovaj skup podataka korišten je u sklopu natjecanja George B. Moody PhysioNet Challenge 2022 (Reyna et al., 2022). U ovom poglavlju bit će opisan taj skup podataka i objašnjena njegova važnost u klasifikaciji srčanih oboljenja.

2.1. Važnost dobrog skupa podataka u fonokardiografiji

Važan i temeljan korak u razvoju računalnih sustava koji pomažu liječnicima kod odlučivanja i dijagnoze kardiovaskularnih bolesti kod auskultacije srca je prikupljanje velikog skupa podataka srčanih zvukova s komentarima i oznakama. Ti podaci moraju moći ispravno predstaviti i karakterizirati šumove i anomalije kod pacijenata s različitim kardiovaskularnim bolestima. Uzevši u obzir današnje moderne tehnike strojnog i dubokog učenja koje zahtijevaju mnogo podataka, potreba za takvim velikim skupovima posebno je ključna (Oliveira et al., 2022). Autori skupa podataka CirCor DigiScope bili su motivirani za prikupljanje velikog skupa podataka, jer je takvih do sada nedostajalo u području auskultacije srca i fonokardiografije.

Trenutačno dostupni označeni i javni skupovi podataka PCG-a imaju ograničeni opseg, zbog toga što pružaju ograničene informacije u vezi s općom procjenom srčanog zvuka (normalan naspram abnormalnog) ili prisutnosti/odsutnosti abnormalnih zvukova kao što su šumovi. U većini dosadašnjih skupova podataka srčani zvukovi klasificirani su samo na temelju općenitih srčanih stanja bez davanja potpunih objašnjenja specifičnih zvučnih karakteristika koje se susreću u svakoj snimci (Oliveira et al., 2022).

Neki od takvih dosad dostupnih skupova podataka su:

- *Pascal Challenge Database* (Gomes et al., 2013), koji se sastoji od 656 snimki PCG-a od nepoznatog broja pacijenata koji su dijelom prikupljeni preko mobilne aplikacije, a dijelom digitalnim stetoskopom na dječjem kardiološkom odjelu bolnice u gradu Recife u sjeveroistočnom Brazilu.
- *PhysioNet/CinC Challenge 2016 Database* (Liu et al., 2016) bio je dio izazova PhysioNet/CinC 2016. Taj skup podataka sadrži ukupno 2435 zapisa srčanih zvukova

od 1297 pacijenata, snimljenih s četiri različita mjesta auskultacije (aortnog, pulmonalnog, trikuspidnog i mitralnog područja).

- *HSCT-11* (Spadaccini & Beritelli, 2013): Ovaj skup podataka sastoji se od ukupno 412 srčanih zvukova od 206 pacijenata, prikupljenih također s četiri mjesta auskultacije. Nedostaju informacije o zdravstvenom stanju pojedinog ispitanika.
- *Digiscope* (Oliveira et al., 2018): Pedijatrijski skup podataka sastavljen je od 29 srčanih zvukova 29 pacijenata, dobi od šest mjeseci do 17 godina. Zvukovi su snimani samo s mitralne auskultacijske pozicije.

Skup podataka CirCor DigiScope ima za cilj riješiti ograničenja prijašnjih skupova podataka, pružajući tako potrebna sredstva za razvoj novih računalno potpomognutih sustava odlučivanja. To se postiže prikupljanjem velikog skupa zvukova koji nude raznoliku karakterizaciju anomalija i njihovim karakteriziranjem pomoću istih ljestvica i parametara koji se koriste u kliničkoj praksi (Oliveira et al., 2022).

2.2. Kampanja i prikupljanje podataka

Predstavljeni skup podataka prikupljen je u sklopu dvije masovne kampanje, koje se nazivaju kampanje "Caravana do Coração" (Karavan srca), provedene u državi Paraíba u Brazilu između srpnja i kolovoza 2014. te lipnja i srpnja 2015. Prikupljeni skup podataka dostupan je na web stranici PhysioNet, iako je trenutno dostupno samo 70 % skupa podataka zbog toga što je skup podataka bio korišten za natjecanje. Ostalih 30 % skupa podataka još uvijek nije javno objavljeno.

Proučavana populacija uključivala je sudionike koji su se dobrovoljno prijavili za vrijeme istraživanja. U kampanjama je prisustvovalo ukupno 2061 sudionika, a 493 sudionika isključeno je iz skupa podataka zbog neispunjavanja kriterija prihvatljivosti (Oliveira et al., 2022).

Svi su sudionici morali ispuniti sociodemografski upitnik i nakon toga su bili provedeni klinički pregledi i kardiološki pregledi. Kasnije je provedena procjena kvalitete podataka i svi unosi su provjereni zbog pogrešno izmjerenih ili pogrešno unesenih vrijednosti, nedosljednih podataka ili prisutnosti stršćih vrijednosti, te su izostavljeni u slučaju bilo kakve nedosljednosti. Nakon toga je izvršena elektronska auskultacija digitalnim stetoskopom i prikupljeni su audio uzorci s četiri auskultacijske točke opisane u prijašnjim poglavljima (mitralno, trikuspidno, aortno i pulmonalno polje). Sve je uzorke prikupljala ista

osoba tijekom trajanja pregleda, u stvarnom kliničkom okruženju. Dva neovisna kardiološka fiziologa pojedinačno su pregledala i ocijenila sve snimljene PCG audio datoteke za kvalitetu signala (Oliveira et al., 2022).

2.3. Demografski i klinički podaci

Prikupljeni skup podataka uključuje 1568 sudionika, 787 (50,2 %) muškaraca i 781 (49,8 %) žena. Što se tiče dobnih kategorija, 988 je djece (63,0 %), 311 dojenčadi (19,80 %), 127 adolescenata (8,1 %), 9 mladih odraslih (0,6 %), 11 novorođenčadi (0,7 %) i 110 trudnica (8,1 %). Kod 12 bolesnika nema podataka o dobi (0,8 %). Prosječna dob sudionika je $73,4 \pm 0,1$ mjesec, u rasponu od 0,1 do 356,1 mjeseci. Prosječna težina i visina uzorka je 24 ± 15 kg, odnosno 111 ± 29 cm, sa srednjim BMI od 18 ± 7 (Oliveira et al., 2022). Po ovim demografskim podacima se jasno vidi kako se radi o uzorku pretežno pedijatarske populacije. Razlog tome je što su prirodene kardiovaskularne bolesti i bolesti srčanih zalistaka najzastupljenije upravo kod takvih populacija.

Potvrđeno je ukupno 647 pojedinačnih ili višestrukih dijagnoza, od kojih su najčešće jednostavne urođene kardiomiopatije (30,2 %) i stečene kardiomiopatije (3,3 %), a utvrđeno je i 65 (3,9 %) dijagnoza složene urođene kardiomiopatije (Oliveira et al., 2022).

2.4. Zvukovi srca i njihovo označavanje

Prikupljeni skup podataka uključuje ukupan broj od 215.780 srčanih zvukova. Od toga je 103.853 srčanih zvukova (51.945 valova S1 i 51.908 valova S2) iz CC2014 i 111.927 (56.449 valova S1 i 55.478 valova S2) iz CC2015 (Oliveira et al., 2022).

Zvučni signali srca prikupljeni su pomoću stetoskopa Littmann 3200 s ugrađenom tehnologijom DigiScope Collector. Signali su uzorkovani na 4 kHz i sa 16-bitnom rezolucijom i normirani su unutar raspona $[-1, 1]$ (Oliveira et al., 2022).

Šumovi su bili prisutni u 305 ispitanika unutar skupa podataka. Od toga je 294 ispitanika imalo samo sistolički šum, 1 ispitanik samo dijastolički šum, a 9 ispitanika imalo je i sistolički i dijastolički šum (Oliveira et al., 2022).

Stečeni audio uzorci automatski su segmentirani pomoću tri algoritma za obradu zvuka. Ovi algoritmi otkrivaju i identificiraju temeljne srčane zvukove (S1 i S2 zvukove) i njihove odgovarajuće granice. Dva kardiološka fiziologa su neovisno jedan o drugome pregledala

rezultate dobivenih algoritama i na temelju pregleda nalaza odlučili se za točnu oznaku svakog od zapisa PCG-a (Oliveira et al., 2022).

Nazivi zvučnih zapisa i odgovarajućih datoteka s oznakama su u formatu `ABCDE_XY.wav`, gdje je ABCDE numerički identifikator pacijenta, a XY je jedan od sljedećih kodova, koji odgovara mjestu auskultacije gdje je PCG sakupljen na površini tijela. PV odgovara pulmonalnoj zoni, TV odgovara trikuspidnoj zoni, AV odgovara zoni aorte, MV odgovara mitralnoj zoni i konačno PhC za bilo koje drugo mjesto auskultacije (Oliveira et al., 2022). Uz svaki audio zapis postoji i datoteka zaglavlja u formatu `ABCDE_XY.he`, u kojoj se nalaze informacije o zvučnom zapisu poput duljine zapisa, frekvencije uzorkovanja, rezolucijom, itd. Također se uz svaki audio zapis nalazi i datoteka s informacijama o točnom vremenu pojavljivanja određenog srčanog zvuka (mogući su S1 i S2), te vremenski periodi sistole i diastole u formatu `ABCDE_XY.tsv`.

U ovom skupu podataka zabilježeni su razni izvori buke, od zvuka trljanja stetoskopa do zvuka smijeha ili plača u pozadini. To znači da je automatska analiza CirCor DigiScope skupa podataka doista težak zadatak. S druge strane, predloženi skup podataka je reprezentativan uzorak stvarnih okruženja u kojima ovakvi sustavi moraju raditi (Oliveira et al., 2022).

S obzirom na veliki broj sudionika iz kampanja CC2014 i CC2015, skup podataka CirCor DigiScope reprezentativan je uzorak ruralnog i urbanog stanovništva u sjeveroistočnom Brazilu, Paraíbi. Pregledani uzorak populacije općenito pokazuje dobro kliničko stanje, s većinom kliničkih i fizioloških parametara unutar normalnog raspona za dob. Unatoč općenito dobrom kliničkom stanju sudionika, pronađene su razne urođene i stečene bolesti. Skup podataka CirCor DigiScope daleko je najveći javno dostupni skup podataka PCG-a (5282 snimke), sadržavajući snimke prikupljene s više auskultacijskih pozicija na tijelu. Nadalje, kampanja je rezultirala vrlo detaljnim skupom podataka s opisima i klasifikacijama šumova (uključujući vrijeme, visinu, ocjenjivanje, oblik, kvalitetu, mjesto auskultacije, itd.), koji se može koristiti u budućim istraživanjima (Oliveira et al., 2022).

Zbog toga što je najveći, reprezentativan, detaljan, pouzdan i slobodno dostupan, podaci iz ovog skupa koristit će se u ovom istraživanju kao ulazi u modele strojnog i dubokog učenja sa svrhom detekcije srčanih šumova i usporedbe uspješnosti modela na tom zadatku.

Podaci su podijeljeni na skupove za učenje, validaciju i testiranje u omjeru 0.7/0.1/0.2. Podjela je obavljena slučajno pritom pazeći na jednaku zastupljenost klasa u svim

skupovima. Kako ne bi došlo do curenja podataka, osigurano je da se zapisi PCG-a istog pacijenta ne mogu naći u različitim skupovima.

3. Metode strojnog učenja

Budući da je cilj ovog istraživanja usporediti uspješnost različitih metoda strojnog i dubokog učenja u detekciji srčanih šumova iz snimki PCG-a, istraživanje započinje s par temeljnih algoritama klasičnog strojnog učenja. Metode strojnog učenja često se koriste kao sustav za podršku u medicini kod donošenja medicinskih dijagnoza. Korišteni algoritmi su algoritam logističke regresije, naivni bayesov klasifikator, te dva algoritma ansambla: klasifikator slučajnih šuma i algoritam XGBoost.

Potrebno je napomenuti kako u korištenom skupu podataka postoji izrazita nebalansiranost u broju primjeraka pojedinih klasa. Naime, u skupu podataka postoje tri klase ciljne značajke koje se odnose na prisutnost srčanog šuma u snimci PCG-a: klasa „Present“ koja označava da je šum prisutan u snimci, klasa „Absent“ koja označava da u snimci srčani šum nije prisutan i klasa „Unknown“ koja označava da nije sigurno postoji li šum u snimci ili ne. Zastupljenost klase „Absent“ je 74 %, klase „Present“ 19 %, a klase „Unknown“ samo 7 %. Budući da su PCG zapisi koji su označeni s „Unknown“ jako niske zastupljenosti u skupu podataka, oni neće biti razmatrani u istraživanju zbog pretpostavke da će jako otežati klasifikaciju podataka i detekciju šumova. Nadalje, tako označeni podaci ne sadrže informaciju o prisustvu šuma u zapisu pa je to još jedan od razloga za njihovo izbacivanje.

Nakon izbacivanja klase „Unknown“ problem nebalansiranog skupa podataka još uvijek nije riješen jer su primjeri s prisutnim srčanim šumom u puno manjem broju. Zbog toga se može pretpostaviti da će takva neravnoteža znatno otežati izgradnju modela koji će moći jednako dobro odlučivati postoji li srčani šum u zapisu ili ne. Rezultati modela pokazat će je li ova pretpostavka bila pravilna.

3.1. Izlučivanje značajki iz audio zapisa

Kako ovakvi modeli strojnog učenja na svoj ulaz primaju podatke u takozvanom tabličnom obliku, oni ne mogu raditi sa „sirovim“ audio zapisima. Tablični oblik podataka je takav da postoje točno određene značajke koje za svaki primjer imaju neku brojčanu ili kategoričku vrijednost.

Iz audio zapisa kao što je PCG može se izlučiti mnoge različite značajke koje bi mogle biti korisne za opisivanje tog signala. Značajke zvuka mogu se podijeliti na dvije najpoznatije grupe: značajke iz vremenske i značajke iz frekvencijske domene.

U vremenskoj domeni često se koriste značajke poput: amplitude (magnituda audio signala u određenom vremenskom trenutku), duljine signala, srednjeg kvadratnog korijena vrijednosti (RMS, daje informaciju o energiji signala), stope prelaska nule (pokazatelj visine tona i buke), entropije signala (mjeri slučajnost u signalu), autokorelacije (sličnost signala i njegove verzije s vremenskim pomakom).

Unatoč mnogim značajkama iz vremenske domene ipak se češće koriste značajke frekvencijske domene, jer one ponekad sadrže dublje i korisnije informacije o signalu. Takve značajke se nazivaju i spektralnim značajkama, jer je naziv postupaka za njihovo dobivanje spektralna analiza. Spektralna analiza definira se kao određivanje frekvencijskog sadržaja valnog oblika. Ova je tehnika uvedena kako bi pružila dublji prikaz signala od onoga što bi se moglo postići jednostavnim tehnikama analize u vremenskoj domeni (Subasi, 2019.).

Za ulaze u modele strojnog učenja koristit će se kepralni koeficijenti mel-frekvencije (engl. *mel-frequency cepstral coefficients*, MFCCs) koji su značajke frekvencijske domene temeljene na diskretnoj Fourierovoj transformaciji. Te su značajke često korištene u analizi zvuka.

3.1.1. Fourierova transformacija

Spektralna analiza temeljena je na metodi Fourierove transformacije (FT). Ona proizvodi skup koeficijenata koji karakteriziraju spektar snage i koreliraju signal sa sinusima i kosinusima različitih frekvencija. Spektar snage može se koristiti za jednostavno dobivanje snage određenog frekvencijskog pojasa. Za točno određivanje spektra, signal mora biti periodičan (ili ograničene duljine) i bez šuma. Međutim, u mnogim biomedicinskim primjenama, signal je beskonačan ili je jednostavno samo njegov dio dostupan za analizu. Također, valni oblik često sadrži značajne količine artefakata i šuma zajedno sa signalom. Uzimajući u obzir ova dva ograničenja, očito je da su sve metode spektralne analize približne i daju procjene pravog spektra. Stoga različite metode spektralne analize pokušavaju poboljšati točnost procjene pojedinih spektralnih značajki (Subasi, 2019).

Klasična kontinuirana FT jedna je od najpopularnijih metoda među brojnim trenutno dostupnim metodama spektralne procjene. FT se temelji na činjenici da sinusne i kosinusne

funkcije imaju energiju samo na jednoj frekvenciji. Glavna ideja je rastaviti valni oblik u niz sinusoida različitih frekvencija. Frekvencijska komponenta valnog oblika na tim frekvencijama je tada izravno proporcionalna amplitudi ovih sinusoida. (Semmlow, 2004).

Bilo koji periodični valni oblik može se predstaviti nizom sinusoida koje su na istoj frekvenciji ili višekratnici frekvencije tog valnog oblika. Svaka sinusoida može se izraziti ili kao sinus i kosinus s odgovarajućim amplitudama, ili kao jedan sinusni val odgovarajuće amplitude i faznog kuta. Kako bi se pronašle odgovarajuće amplitude ovih sinusoida, potrebno je pomnožiti valni oblik s funkcijom sinusa i kosinusa i integrirati valni oblik po vremenu. Sljedeće dvije jednadžbe (3.1) i (3.2) objašnjavaju ovaj proces (Semmlow, 2004):

$$a(m) = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \cos(2\pi m f_T t) dt \quad (3.1)$$

$$b(m) = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \sin(2\pi m f_T t) dt \quad (3.2)$$

gdje je T period ili duljina valnog oblika, $f_T = 1/T$, a m je cijeli broj ($m = 1, 2, 3 \dots$).

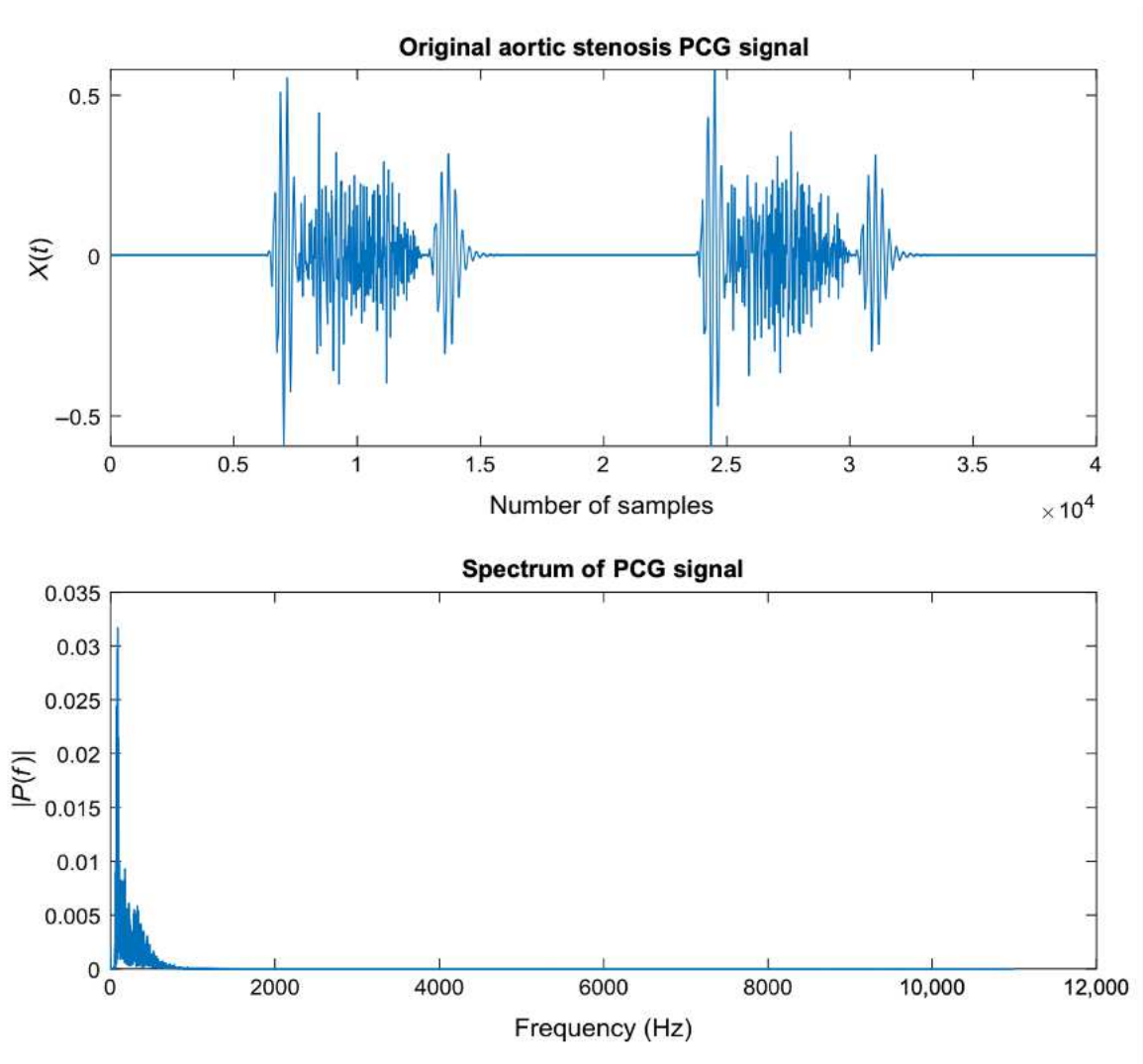
Diskretna Fourierova analiza proširenje je kontinuirane Fourierove analize. Kod diskretne FT uvode se dvije dodatne stavke: uzorkovanje i prozori. Ove operacije uvedene su zbog toga što se signali moraju spremati u stvarnu memoriju na računalo i nemogu biti beskonačni. Tako će biti potrebno odabrati frekvenciju uzorkovanja i oblik i veličinu prozora u koji se signal smješta. Teorem uzorkovanja (Nyquist) kaže da analogni valni oblik mora biti frekvencijski ograničen na način da je njegova najviša frekvencija manja od polovine Nyquistove frekvencije. To znači da frekvencija uzorkovanja mora biti veća ili jednaka Nyquistovoj frekvenciji. Na taj se način postiže da digitalizirani podaci nose sve relevantne informacije izvornog analognog valnog oblika. Operacija prozora (engl. *windowing*) je proces u kojem digitalizirani valni oblici moraju nužno biti skraćeni zbog skladištenja i drugih ograničenja. Ovaj proces se također može smatrati množenjem podataka s nekim oblikom prozora. Najčešći oblik prozora je pravokutni prozor, čijim množenjem signal ne mijenja svoj oblik (Subasi, 2019). Jednadžbe za izračunavanje analize Fourierovog niza digitaliziranih podataka iste su kao za kontinuirane podatke osim što je integracija zamijenjena zbrajanjem, a sinusoida zapisane koristeći Eulerov identitet kao što je prikazano jednadžbom (3.3) (Semmlow, 2004).

$$e^{jx} = \cos x + j \sin x \quad (3.3)$$

Izraz za diskretnu Fourierovu transformaciju sada postaje (3.4) (Semmlow, 2004):

$$X(m) = \sum_0^{N-1} x(n)e^{-\frac{j2\pi mn}{N}} \quad (3.4)$$

Slika 3.1 Fourierova transformacija PCG signala prikazuje jednu Fourierovu transformaciju PCG signala, iz koje se vidi da je frekvencijski raspon PCG-a dosta nizak.



Slika 3.1 Fourierova transformacija PCG signala (Subasi, 2019.)

3.1.2. Kepstralni koeficijenti mel-frekvencije

Konačno dolazimo do kepstralnih koeficijenata mel-frekvencije (MFCC) koji su bazirani na diskretnoj FT.

Kepstralni koeficijenti mel-frekvencije vrlo su popularan način izlučivanja značajki u području obrade zvuka. MFCC-ovi su zapravo tip kepstralne reprezentacije signala, gdje frekvencijski pojasevi nisu raspoređeni linearno, već prema mel-ljestvici. Kako bi iz određenog vremenskog okvira signala dobili značajke MFCC, potrebno je nekoliko koraka (Giannakopoulos & Pirkakis, 2014).

Prvi korak je računanje diskretne FT (Giannakopoulos & Pirkakis, 2014).

Drugi korak je dobiveni spektar provesti kroz L funkcija filtra na mel-skali. Mel-skala uvodi efekt zakrivljenja frekvencije sa svrhom da se uskladi s određenim psihoakustičkim opažanjima. Naime utvrđeno je da ljudski slušni sustav može lakše razlikovati susjedne frekvencije u niskofrekventnom području. Zapravo, mel-ljestvica je ljestvica frekvencijskih intervala koja je konstruirana kako bi odgovarala ljudskoj percepciji. Ako ih procijeni ljudski slušatelj, frekvencijski intervali djelovat će kao da su jednako razmaknuti (Giannakopoulos & Pirkakis, 2014).

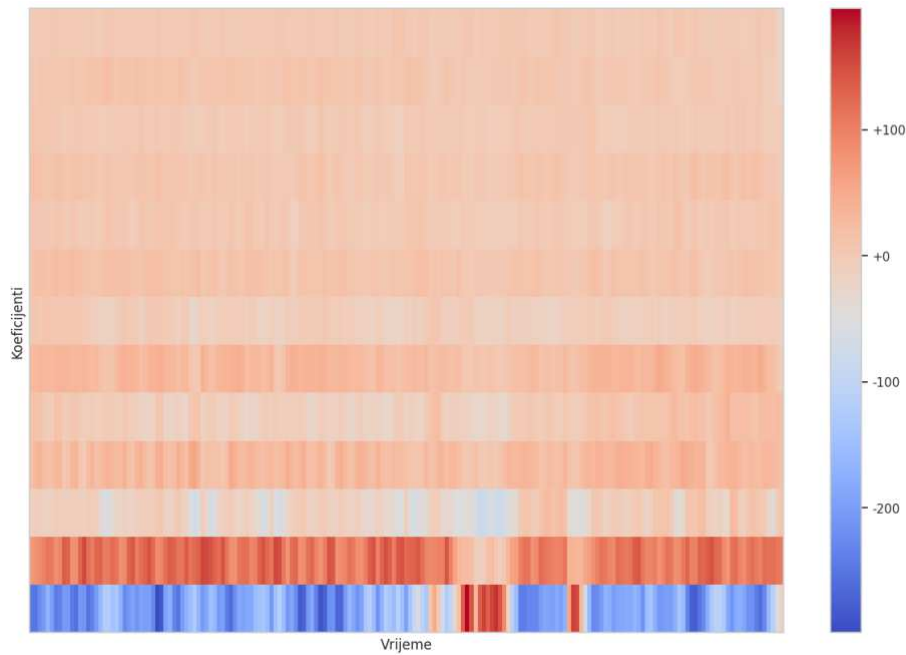
U trećem koraku se računa MFCC iz jednadžbe (3.5). Gdje je \widetilde{O}_k , $k = 1, \dots, L$, snaga na izlazu k -tog filtra (Giannakopoulos & Pirkakis, 2014).

$$t_m = \sum_{k=1}^L (\log \widetilde{O}_k) \cos \left[m \left(k - \frac{1}{2} \right) \frac{\pi}{L} \right] \quad (3.5)$$

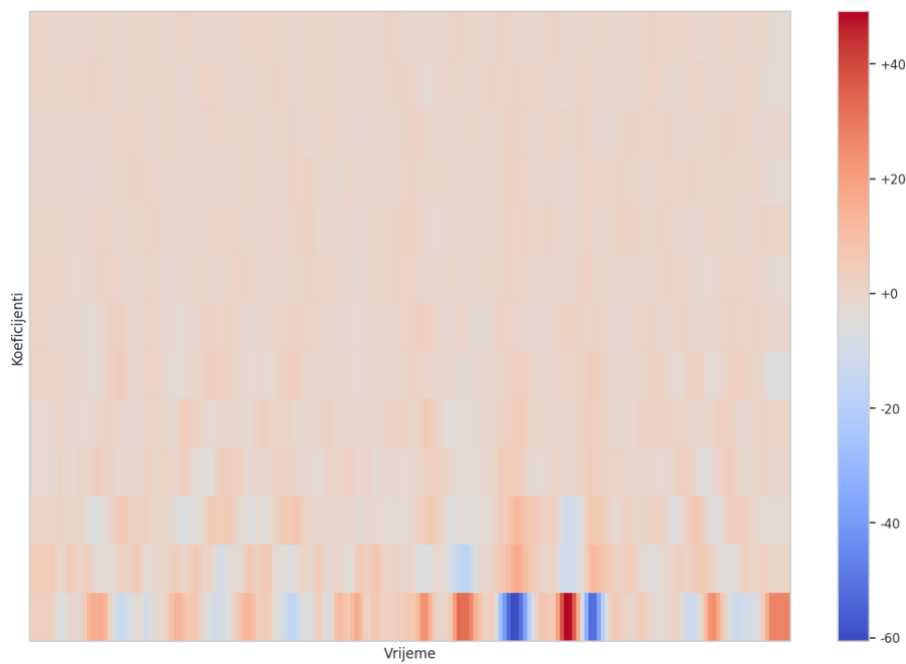
Stoga, prema jednadžbi (3.5), MFCC su diskretni koeficijenti kosinusne transformacije logaritamskog spektra snage na mel-skali. MFCC se naširoko koriste u prepoznavanju govora, klasifikaciji glazbenih žanrova, grupiranju govornika i mnogim drugim primjenama audio analize (Giannakopoulos & Pirkakis, 2014). Rezultati ovog istraživanja će pokazati jesu li dovoljno informativni za razlikovanje PCG signala.

U mnogim aplikacijama za obradu glazbe postalo je uobičajeno odabrati prvih 13 MFCC-ova jer se smatra da nose dovoljno diskriminirajućih informacija u kontekstu raznih zadataka klasifikacije (Giannakopoulos & Pirkakis, 2014). Po uzoru na to, za ulaze u korištene klasifikatore uzima se prvih 13 MFCC-a, kao i njihove prve i druge derivacije, što sve zajedno čini 39 značajki po svakom zapisu PCG-a.

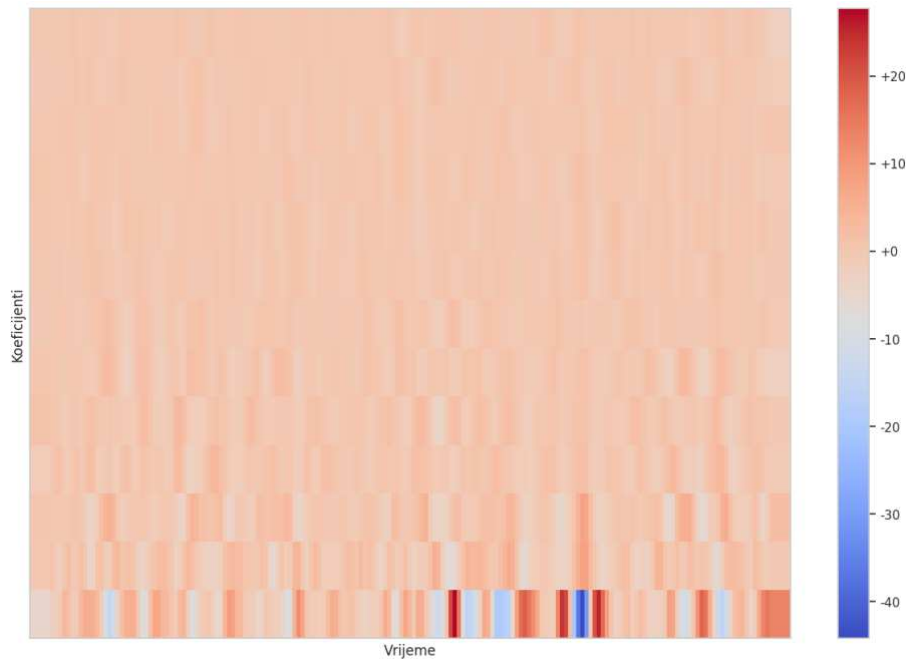
Slika 3.2 prikazuje prvih 13 značajki MFCC jednog zapisa PCG-a, dok Slika 3.3 i Slika 3.4 prikazuju prvu i drugu derivaciju tih značajki. Toplije boje odgovaraju većim vrijednostima, dok hladnije boje na grafu odgovaraju nižim vrijednostima koeficijenata.



Slika 3.2 Grafički prikaz prvih 13 MFCC-a jednog PCG zapisa



Slika 3.3 Grafički prikaz derivacije prvih 13 MFCC-a sa Slike 3.2



Slika 3.4 Grafički prikaz druge derivacije MFCC-a sa Slike 3.2

3.2. Logistička regresija

Prvi od klasičnih modela strojnog učenja koji je bio ispitan i naučen je model logističke regresije.

Logistička regresija je algoritam klasifikacije koji je vrlo jednostavno implementirati, ali svejedno daje vrlo dobre rezultate na linearno odvojivim problemima. To je jedan od najčešće korištenih algoritama za klasifikaciju u industriji (Raschka, 2015).

Za razliku od linearne regresije gdje se srednja kvadratna pogreška (engl. *mean squared error*, MSE) ili njezin korijen (RMSE) koristi kao funkcija gubitka, logistička regresija koristi funkciju gubitka koja se naziva procjena maksimalne izglednosti (engl. *maximum likelihood estimator*, MLE) koja je uvjetna vjerojatnost. Ako je vjerojatnost veća od 0.5, primjer će se klasificirati u prvu klasu, a u suprotnom bit će klasificiran u drugu klasu (Belyadi & Haghghat, 2021).

Bazna struktura modela logističke regresije prikazana je formulom (3.6).

$$z = w_0x_0 + w_1x_1 + \dots + w_mx_m = \sum_{i=0}^m w_ix_i = \mathbf{w}^T \mathbf{x} \quad (3.6)$$

Logistička regresija na svom izlazu koristi sigmoidnu funkciju koja se često naziva i logističkom funkcijom. Ona je prikazana formulom (3.7).

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (3.7)$$

Tako je formula za model logističke regresije (3.8):

$$h(x; w) = \sigma(w^T x) = \frac{1}{1 + e^{-(w^T x)}} \quad (3.8)$$

Rezultati ovog i svih ostalih modela strojnog i dubokog učenja bit će vrednovani kroz više standardnih metrika kao što su točnost modela (engl. *accuracy*), preciznost (engl. *precision*), odziv ili osjetljivost (engl. *recall*, *sensitivity*), i F1 mjera (engl. *F1 score*). Svaka od ovih metrika računa se preko rezultata klasifikacije koji se iznose u obliku točno klasificiranih pozitivnih primjeraka (engl. *true positives* - TP), lažno pozitivnih primjeraka (engl. *false positives* - FP), točno klasificiranih negativnih primjeraka (engl. *true negatives* - TN) i lažno negativnih primjeraka (engl. *false negatives* - FN). U slučaju ovog ekperimenta, lažno pozitivan primjerak je onaj koji je klasificiran tako da je u njemu detektiran srčani šum, a zapravo ga nema.

$$\text{točnost} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.9)$$

$$\text{preciznost} = \frac{TP}{TP + FP} \quad (3.10)$$

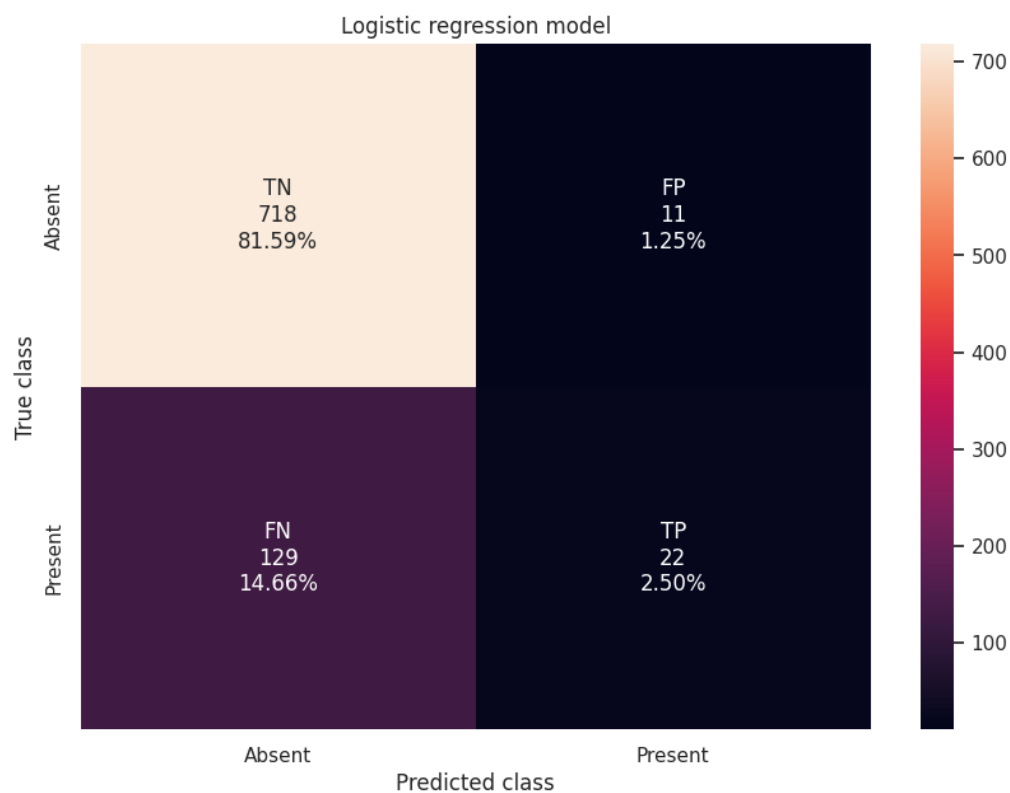
$$\text{odziv} = \frac{TP}{TP + FN} \quad (3.11)$$

$$\text{f1 mjera} = 2 \frac{\text{preciznost} * \text{odziv}}{\text{preciznost} + \text{odziv}} = \frac{2TP}{2TP + FP + FN} \quad (3.12)$$

Preciznost je prikladna metrika kada je cijena lažno pozitivnog rezultata vrlo visoka, a obrnuto vrijedi za odziv koji je važan u slučaju kada je cijena lažno negativnog rezultata vrlo visoka. U slučaju detekcije srčanih šumova kod PCG-a, metrika koju je važno promatrati je upravo odziv, jer je bitnije donijeti točnu dijagnozu za zapise u kojima postoji srčani šum, da ne bi prošli neopaženo. F1 mjera je također bitna metrika jer uzima u obzir i preciznost i odziv, na taj način daje realan rezultat klasifikacije modela u slučaju nebalansiranosti klasa, za razliku od točnosti koja u tom slučaju nije realan pokazatelj uspješnosti. Točnost je loša metrika kod nebalansiranih skupova podataka jer bi pristrani algoritam mogao klasificirati većinu primjera kao većinsku klasu što bi se prikazalo kao visoka točnost.

Što se tiče hiperparametara, ovaj algoritam ih ima nekoliko. Postoji hiperparametar *penalty* koji odabire način regularizacije modela. Regularizacija je metoda kojom se reguliraju veličine težina modela, pa se tako postiže bolja generalizacija i smanjuje prenaučenosť. Korištena je regularizacija L2 kao najpopularnija opcija. Uz regularizaciju se veže i hiperparametar *C* koji je inverz regularizacijskog multiplikatora *lambda*. Što je *C* veći, to će biti manje regularizacije težina. Korištena je vrijednost $C = 1$, jer takvim odabirom regularizacija je od srednje važnosti. Hiperparametar *max_iter* postavljen je na 100. On označava maksimalan broj iteracija algoritma optimizacije.

Na Sliku 3.5 prikazana je matrica zabune za model logističke regresije.



Slika 3.5 Matrica zabune za model logističke regresije

Ovaj model na skupu za testiranje postigao je točnost od 84 %, preciznost od 67 %, odziv od 16 % te F1 mjeru od 24 %. Po ovim se rezultatima vidi kako točnost kod ovog skupa podataka nije validna metrika, jer je po vrijednosti visoka, ali model ne razlučuje dobro klase, te je za 129 primjera pogrešno zaključio da nemaju srčani šum.

3.3. Naivni Bayesov klasifikator

Naivni Bayesov klasifikator verzija je Gaussovog Bayesovog klasifikatora. Taj model koristi vjerojatnosti kako bi donio odluku o ciljnoj varijabli nekog primjera. Baza ovog modela je Bayesov teorem, koji opisuje vjerojatnost događaja na temelju prethodnog znanja o uvjetima koji se odnose na događaj. Model Gaussovog Bayesovog klasifikatora prikazan je formulom (3.13).

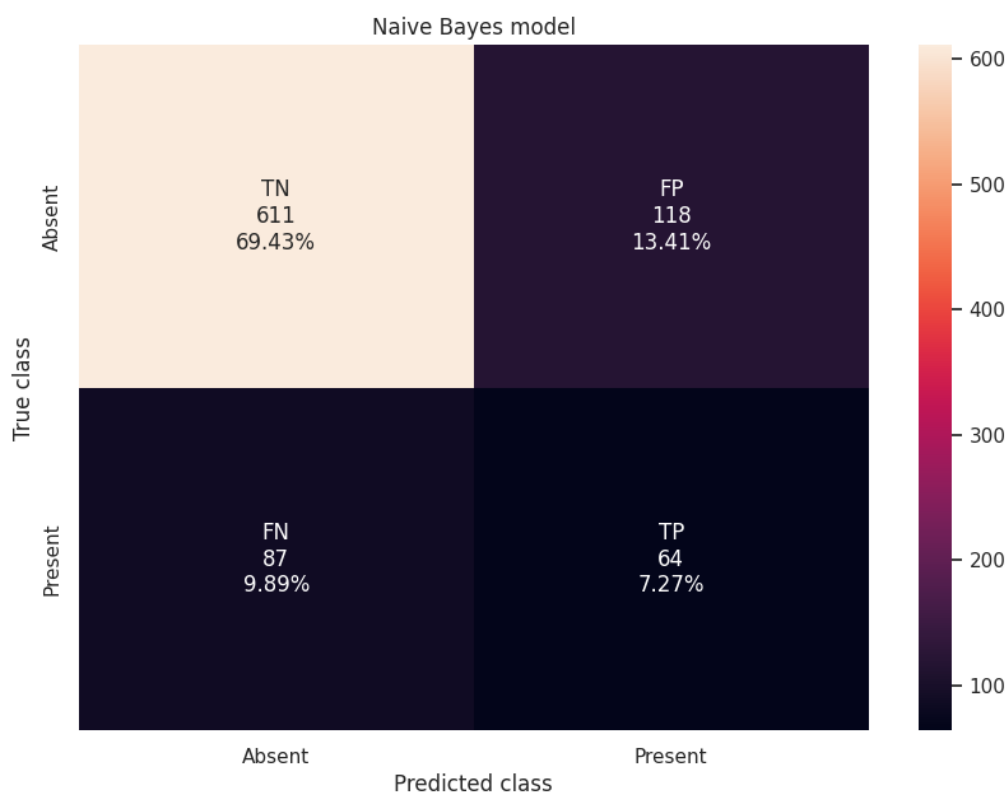
$$h(x) = \operatorname{argmax}_y p(x|y)P(y) \quad (3.13)$$

Po formuli (3.13), neki primjer x klasificira se u određenu klasu y čija je vjerojatnost najveća. $P(y)$ označava apriornu vjerojatnost klase, dok $p(x|y)$ označava gustoću vjerojatnosti primjera za zadanu klasu. Gustoća vjerojatnosti primjera za zadanu klasu govori kolika je vjerojatnost ako se gleda određena klasa y , da se u njoj nađe primjer x . Budući da je ovo Gaussov Bayesov klasifikator, za izračun gustoće vjerojatnosti $p(x|y)$ koristi se normalna (Gaussova) razdioba. Inače, za točnu vjerojatnost klase y u formuli (3.13) bio bi potreban dodatan nazivnik koji je sastavni dio Bayesovog teorema, ali budući da se vjerojatnost za svaku klasu računa s tim istim nazivnikom, za pronalazak najveće vjerojatnosti on u formuli nije potreban.

Ako se uvede pretpostavka da su sve značajke međusobno uvjetno nezavisne uz zadanu klasu y , tada se dobiva model naivnog Bayesovog klasifikatora.

Na početku je potrebno eliminirati značajke koje su međusobno visoko linearno korelirane, jer će postupkom razdvajanja klasa biti teško razlučiti između njih. Pretpostavka da su interakcije jednake nuli dovodi do implicitne regularizacije, tako da je varijanca modela smanjena, što rezultira klasifikacijama koje su točnije u mnogim primjenama. Površina odluke naivnog Bayesovog klasifikatora složenog je i nelinearnog oblika. Kao algoritam, lagan je za implementaciju, jer ne zahtijeva kompliciranije iterativne postupke optimizacije za estimaciju parametara. Ovaj model nije uvijek najbolji mogući klasifikator za neke zadatke, međutim, može biti prilično dobar pod određenim okolnostima (Wu et al., 2008; Subasi, 2020).

Matrica zabune naivnog Bayesovog klasifikatora na problemu detekcije srčanih šumova prikazana je Slika 3.6.



Slika 3.6 Matrica zabune za model naivnog Bayesa

Model naivnog Bayesovog klasifikatora na skupu za testiranje postigao je točnost od 77 %, preciznost od 35 %, odziv od 42 % i F1 mjeru od 38 %. U rezultatima se vidi napredak naspram modela logističke regresije koji je klasificirao skoro sve primjere u većinsku klasu, dok je naivni Bayesov klasifikator koristeći vjerojatnosti klasificirao čak 20,68 % primjera u manjinsku klasu, što otprilike odgovara njenoj zastupljenosti u skupu podataka. Naime, ovakav napredak svejedno nije zadovoljavajuć, jer vjerojatnosti nisu dovoljne da riješe problem detekcije srčanih šumova.

3.4. Ansambl slučajna šuma

Model slučajne šume također je jedan od poznatih i često korištenih modela za problem klasifikacije. On je usko povezan s algoritmima za izgradnju stabla odluke (engl. *decision tree*).

Prije objašnjenja slučajne šume, bitno je razumjeti kako funkcionira stablo odluke. Stablo odluke dijeli podatke u podstabla koja se zatim dijele dalje na druga podstabla. Na najvišoj razini nalazi se čvor korijen koji predstavlja cjelokupnu populaciju značajki, čvorovi odlučivanja nalaze se u nižim razinama stabla i imaju dvije ili više grana, i na kraju na

posljednjoj razini nalaze se čvorovi listovi koji se dalje ne dijele, već donose odluke o ciljnoj varijabli za određeni podatak (Belyadi & Haghghat, 2021).

Poznati su mnogi algoritmi stabla odluke kao što su ID3, C4.5, C5.0 i CART. ID3 je kratica za „Iterative Dichotomiser 3“ koji je Quinlan razvio 1986. (Quinlan, 1986.). Ovaj algoritam gradi stabla odlučivanja koristeći pohlepnu heuristiku odozgo prema dolje na kategoričkim značajkama uzimajući u obzir najveću informacijsku dobit (engl. *information gain*) (Belyadi & Haghghat, 2021).

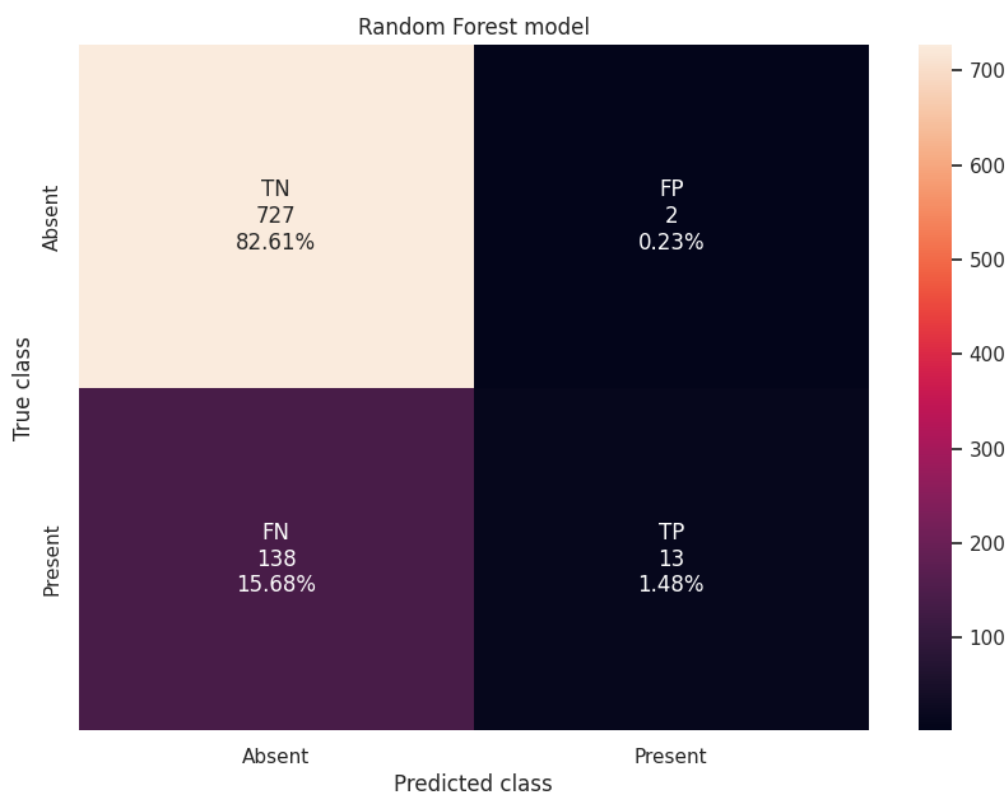
Stablo odluke gradi se dijeljenjem značajki po čvorovima stabla. Odluka u kojem čvoru bi trebala biti koja značajka može biti složena i izazovna. Postoje kriteriji po kojima se to određuje kao što su entropija i informacijska dobit. Entropija je mjera nesigurnosti ili nereda u skupu podataka. Informacijska dobit daje informaciju koliko dobro pojedina značajka odvaja primjere prema njihovoj ciljnoj klasi. Entropija i informacijska dobit obrnuto su proporcionalni pojmovi.

Ansambl je skup od više modela strojnog učenja koji se koristi s ciljem poboljšanja rezultata koje daju pojedinačni modeli. Slučajna šuma ansambl je koji koristi velik broj stabala odluke. Klasifikacija određenog primjera se donosi agregacijom rezultata klasifikacije pojedinačnih stabala većinskim glasanjem. Poznato je da stabla odluke imaju veliku varijancu, pa se ansamblom slučajne šume želi postići smanjenje te varijance. Smanjenje varijance postiže se uglavnom velikim brojem stabala i slučajnim odabirom značajki (iz određenog podskupa značajki) u svakom čvoru.

Predikcija napravljena jednim stablom odluke može biti netočna, ali u kombinaciji s više stabala, predikcija će biti bliže prosjeku. Razlog zašto je slučajna šuma obično točnija od jednog stabla odluke je taj što koristi mnogo predikcija pa se samim time raspolaže s više znanja (Belyadi & Haghghat, 2021).

Ovaj model zahtijeva nekoliko hiperparametara. Broj stabala odluke postavljen je na 100. Maksimalna dubina stabla nije postavljena jer bi to rezultiralo podrezivanjem stabala. Od ostalih hiperparametara izabrani su oni uobičajeni.

Iz matrice zabune na Slika 3.7 vidi se da model slučajne šume nije bio dobar odabir kod detekcije srčanih šumova. Naime, njegova mana je osjetljivost na nebalansiranost klasa, koja je uzrokovala veliku pristranost modela, jer je model većinu primjeraka svrstao u većinsku klasu.



Slika 3.7 Matrica zabune za model slučajne šume

Klasifikator slučajne šume na postigao je točnost od 84 %, preciznost od 87 %, odziv od 7 % i F1 mjeru od 16 % na skupu za testiranje. Iako su točnost i preciznost visoke, opet je bitno napomenuti kako to u ovom slučaju nisu dobre metrike za vrednovanje ovog modela, jer su odziv i F1 mjera dosta niske. Zaključuje se kako slučajna šuma nije dala zadovoljavajuće rezultate, zbog nemogućnosti odvajanja klasa.

3.5. XGBoost

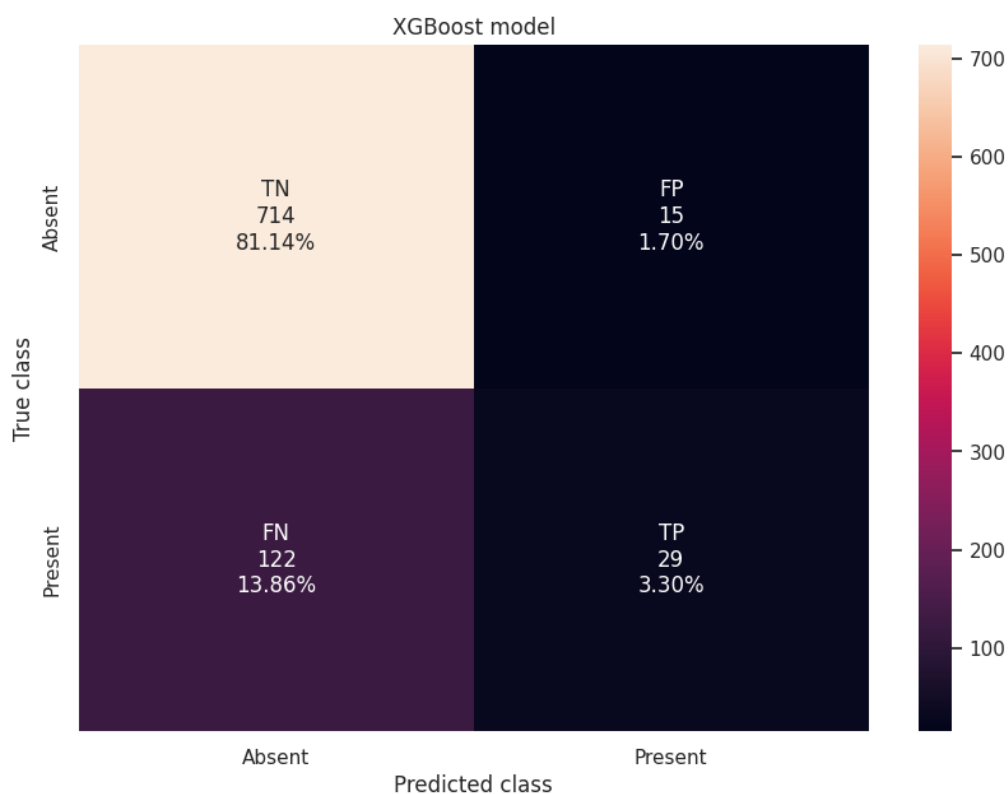
Algoritam eXtreme Gradient Boosting predstavio je (Chen & Guestrin, 2016).

XGBoost je također ansambl modela, ali drugačiji od slučajne šume. Naime, slučajna šuma gradi se od velikog broja pojedinačnih modela, dok XGBoost koristi „boosting“ metodu izgradnje ansambla koja je iterativna i temelji se na optimizaciji funkcije gubitka algoritmom sličnim gradijentnom spustu. „Boosting“ kreće od početnog modela strojnog učenja koji se mijenja u svakoj iteraciji na način da se u model dodaju nova stabla odluke tipa CART (engl. *classification and regression trees*), a primjerima iz skupa podataka koji su bili loše klasificirani dodaju se veće težine (Jović, 2023).

Najvažniji čimbenik uspjeha XGBoosta je njegova skalabilnost u mnogim scenarijima. Sustav radi puno brže od mnogih postojećih popularnih rješenja na jednom računalu i skalabilan je u distribuiranim sustavima i sustavima s ograničenom memorijom (Chen & Guestrin, 2016).

Ovaj algoritam nudi mnogo hiperparametara za prilagođavanje, ali odabrani su oni uobičajeni, jer su takvi hiperparametri zadani optimizacijom ovog algoritma u mnogim primjenama.

XGBoost na skupu za testiranje rezultata klasifikacije PCG zapisa postigao je točnost od 84 %, preciznost od 66 %, odziv od 19 % i F1 mjeru od 30 %. Matrica zabune dana je na Slika 3.8. Iako se rezultati nisu znatno poboljšali od logističke regresije i slučajne šume, ipak je XGBoost algoritam dao malo bolje rezultate, što potvrđuje zašto je jedan od najčešće korištenih u strojnom učenju.



Slika 3.8 Matrica zabune modela XGBoost

Sljedeća tablica (Tablica 1) prikazuje rezultate ispitanih modela klasičnog strojnog učenja.

Tablica 1 Rezultati algoritama strojnog učenja

| <i>Algoritam</i> | Točnost | Preciznost | Odziv | F1 mjera |
|------------------------------------|----------------|-------------------|--------------|-----------------|
| <i>Logistička regresija</i> | 84,10 % | 66,67 % | 14,57 % | 23,91 % |
| <i>Naivni Bayesov klasifikator</i> | 76,70 % | 35,16 % | 42,38 % | 38,44 % |
| <i>Slučajna šuma</i> | 84,10 % | 86,67 % | 8,61 % | 15,66 % |
| <i>XGBoost</i> | 84,43 % | 65,91 % | 19,21 % | 29,74 % |

Svi modeli bit će rangirani prvo po F1 mjeri, nakon toga po odzivu, preciznosti i na kraju po točnosti. Iz toga slijedi da je najbolji model Naivni Bayesov klasifikator, a nakon njega slijede XGBoost, logistička regresija i na kraju slučajna šuma.

Algoritmi strojnog učenja kod ovakvih zadataka često trebaju ekspertne značajke koje proizlaze iz znanja domene problema. Budući da MFCC značajke nisu ekspertne iz područja fonokardiografije, već značajke koje mogu opisivati bilo koji audio signal, one nisu pružile dovoljno znanja za detekciju srčanih šumova. To je jedan od uzroka niskih metrika poput F1 mjere i odziva. Kako algoritmi strojnog učenja nisu dali zadovoljavajuće rezultate, sljedeća ispitana grupa modela bit će duboki modeli neuronskih mreža koji bi mogli dati bolje rezultate na ovom problemu.

4. Konvolucijski i povratni modeli

Na redu su modeli dubokog učenja koji su testirani na problemu detekcije srčanih šumova iz snimki PCG-a. U ovom poglavlju bit će opisani modeli konvolucijskih arhitektura i povratnih arhitektura, te njihove kombinacije. Nadalje, bit će izneseni i uspoređeni dobiveni rezultati ovih modela.

4.1. Duboko učenje

Duboko učenje je grana umjetne inteligencije koja pokušava riješiti probleme koji su za ljude lako rješivi, ali ih je teže formalno opisati. To su problemi koje ljudi rješavaju putem intuicije, koji se čine automatski, poput prepoznavanja izgovorenih riječi ili lica na slikama (Bengio et al., 2017).

Modeli dubokog učenja su slijed naučenih nelinearnih transformacija. Duboke mreže strukturno su slojevite, što znači da sadržavaju više slojeva funkcija. Svaki sloj duboke neuronske mreže sastoji se od jedne ili više linearnih funkcija koje se na izlazu iz sloja propagiraju kroz aktivacijsku funkciju koja ima ulogu nelinearne transformacije. Ovakvi slojevi nastali su po uzoru na neurone u ljudskom mozgu, gdje su linearne operacije veze između više neurona, a nelinearne operacije su aktivacijske funkcije neurona.

Svaki sloj ima određene parametre koji se metodama optimizacije kroz učenje na skupu primjera prilagođavaju prema vrijednostima koje daju najbolju točnost kod klasifikacije. To se postiže određivanjem odgovarajuće funkcije gubitka kojom se daje mjera uspješnosti modela u fazi učenja. Kod optimizacije modela, računaju se gradijenti funkcije gubitka i gradijenti svih parametara mreže koji se onda popravljaju i namještaju u smjeru negativnog gradijenta, jer je to smjer u kojem će model dati bolju točnost. Svaki algoritam dubokog učenja mora sadržavati arhitekturu mreže, odgovarajuću funkciju gubitka i algoritam optimizacije.

Modeli dubokog učenja pokazuju bolje rezultate na velikim skupovima podataka, jer njihovo učenje iziskuje puno informacija kako bi oni konvergirali u minimum funkcije pogreške. Također su uspješni s kompozitnim podacima koji u sebi sadrže neke djeljive komponente. To proizlazi iz načina na koji duboki model računa predikciju. Naime, podatkovna reprezentacija u određenom sloju gradi se od reprezentacija prijašnjeg sloja, i tako svaki sloj

razlučuje podatke na specifičnije komponente i mijenja reprezentaciju podataka. Biomedicinski signali imaju to svojstvo pa su duboki modeli odličan izbor za njihovu klasifikaciju.

4.2. Konvolucijski modeli

Konvolucijski modeli spadaju u modele dubokog učenja. Iako se najčešće koriste u području računalnog vida i obrade slike, također se pokazalo da konvolucijske mreže rade učinkovito s biomedicinskim signalima.

Konvolucijski slojevi nazvani su po operaciji konvolucije koju provode nad podacima. Za razliku od potpuno povezanih slojeva gdje je svaki neuron povezan sa svakim neuronom iz prošlog sloja, konvolucijski slojevi koriste takozvani filter ili jezgru (engl. *kernel*) koja „klizi“ po podacima najčešće s lijeva nadesno i obavlja operaciju konvolucije. Operacija konvolucije je zapravo skalarni umnožak elemenata jezgre s određenim elementima prošlog sloja neurona. Izlaz iz određenog konvolucijskog sloja su rezultati svih operacija konvolucije dobivenih pomicanjem jezgre po podacima. Izlaz iz konvolucije naziva se karta značajki (engl. *feature map*), jer se takva operacija može shvatiti kao izlučivanje značajki iz podataka.

Postoji više vrsta konvolucije s obzirom na dimenzionalnost podataka. Kad su podaci jednodimenzionalni, koristi se jednodimenzionalna jezgra koja klizi s lijeva nadesno. Ako su podaci dvodimenzionalni, onda se koristi dvodimenzionalna jezgra najčešće kvadratnog oblika koja klizi s lijeva nadesno i odozgo prema dolje. Moguća je i trodimenzionalna konvolucija koja onda koristi trodimenzionalne jezgre.

Kako bi se ubrzalo učenje modela i održala stabilnost procesa učenja, kod konvolucijskih modela u praksi se često koriste slojevi normalizacije po grupi (engl. *batch normalization*) koji normiraju svaku grupu podataka. Kad ne bi bilo normalizacije po grupi, moglo bi doći do pomaka distribucije u izlazu pojedinog sloja mreže, pa bi sljedeći sloj osim učenja klasifikacije, morao reagirati i prilagoditi se na takav pomak distribucije. To znatno usporava učenje mreže, a veličina problema proporcionalna je broju slojeva u mreži (Buduma et al., 2022).

Na izlazu iz konvolucije često se koriste slojevi sažimanja (engl. *pooling layers*) koji po nekom principu sužavaju reprezentaciju dobivenih podataka s ciljem da ostanu one značajke koje su najrelevantnije. Postoje razne varijante slojeva sažimanja, a često korišteni su

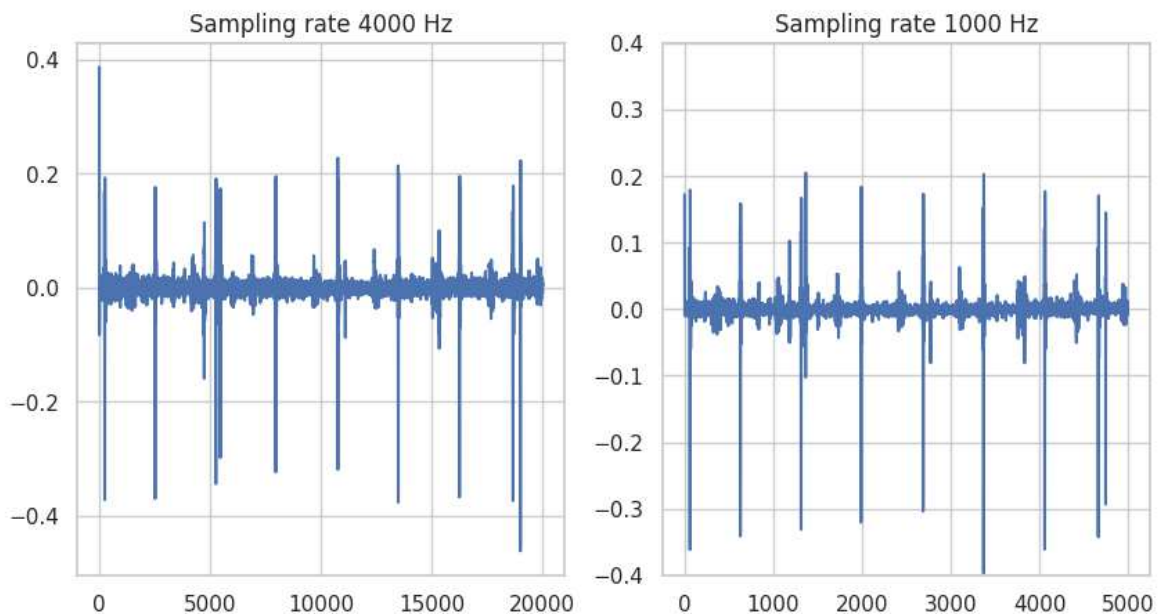
sažimanje najvećom vrijednošću (engl. *max pooling*), ili sažimanje srednjom vrijednošću (engl. *average pooling*).

Podaci na izlazu konvolucijskih dubokih modela ovise o lokalnom susjedstvu podataka na ulazu, pa tako u praksi konvolucija može detektirati razne oblike i anomalije kod biomedicinskih signala. Korisno svojstvo ovakvih mreža je invarijantnost na pomak, što znači da kada se u signalu pojavi značajka koja ukazuje na svojstvo neke od klasa, konvolucija to prepoznaje bez obzira na lokaciju te značajke u signalu. Invarijantnost na pomak korisna je kod problema gdje je bitnije detektirati prisutnost određene značajke, nego njezinu točnu lokaciju, kao što je slučaj kod detekcije srčanih šumova.

4.2.1. Jednodimenzionalni konvolucijski model

Prvi model neuronske mreže koji je ispitan bio je jednodimenzionalni konvolucijski model. Budući da ovaj model koristi jednodimenzionalnu konvoluciju, na njegov ulaz šalju se audio zapisi PCG-a. Audio zapis PCG-a je po strukturi jednodimenzionalan vremenski niz koji ima svoju frekvenciju uzorkovanja.

Kako su svi PCG zapisi u korištenom skupu podataka uzorkovani na frekvenciji 4000 Hz, a informativni frekvencijski pojas kod PCG-a je puno manji od toga (kao što je vidljivo na Slika 3.1), svaki primjer poduzorkovan je na 1000 Hz prije nego je korišten u modelu. Na taj način postiže se smanjenje memorijskog zauzeća podataka, smanjenje vremena učenja modela, a moguće i poboljšanje rezultata. Na primjeru jednog PCG-a bez srčanog šuma, prikazanog na Slika 4.1, može se vidjeti da signal nije značajno promijenjen nakon poduzorkovanja makar mu je duljina smanjena 4 puta.



Slika 4.1 Primjer poduzorkovanja PCG-a

Konvolucijski modeli ne mogu raditi s podacima različitih duljina. Zbog toga su svi zapisi dovedeni na jednake duljine. Oni zapisi koji su bili prekratki nadopunjavani su nulama na kraju. To zapravo znači da je dodana tišina na kraj signala, jer nula kod audio zapisa označava da nema zvuka. Oni zapisi koji su bili predugi, skraćeni su od kraja kako bi odgovarali zajedničkoj duljini.

Slika 4.2 prikazuje arhitekturu jednodimenzionalnog konvolucijskog modela. On se sastoji od 2 konvolucijska sloja s jezgrom veličine 5, pomakom jezgre za 1 i nadopunjavanjem podataka na rubovima za 2 broja. Nakon svakog konvolucijskog sloja radi se normalizacija po grupi i sažimanje podataka maksimalnom vrijednošću. Nakon 2 konvolucijska sloja dolaze 2 potpuno povezana sloja neurona koja služe klasifikaciji reprezentacije podataka dobivene konvolucijom.

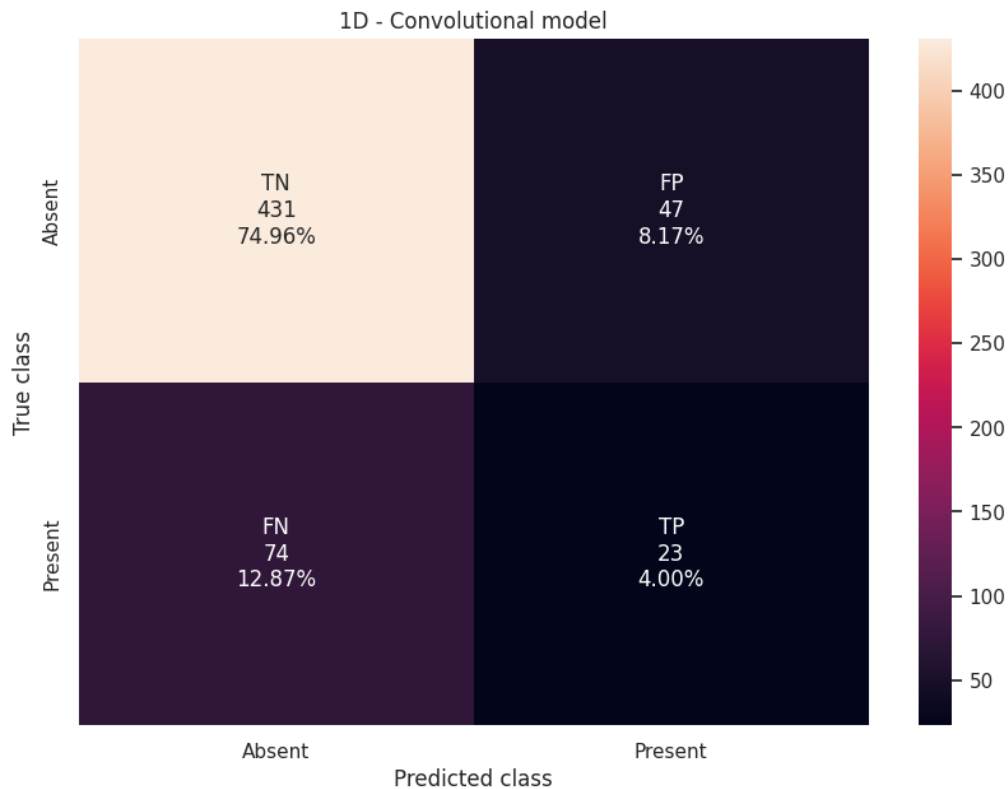
```
ConvolutionalModel(
  (conv1): Conv1d(1, 16, kernel_size=(5,), stride=(1,), padding=(2,))
  (bn1): BatchNorm1d(16, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (maxpool1): MaxPool1d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (conv2): Conv1d(16, 32, kernel_size=(5,), stride=(1,), padding=(2,))
  (bn2): BatchNorm1d(32, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (maxpool2): MaxPool1d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (fc1): Linear(in_features=96000, out_features=512, bias=True)
  (fc2): Linear(in_features=512, out_features=1, bias=True)
)
```

Slika 4.2 Arhitektura jednodimenzionalnog konvolucijskog modela

Za optimizaciju modela korišten je poznati i često korišteni algoritam Adam, sa stopom učenja (engl. *learning rate*) 0.0001 i smanjenjem težina (engl. *weight decay*) 0.00001. Adam

je dosta robusan algoritam koji dobro radi sa širokim skupom parametara. Veličina grupe podataka koji su bili dani modelu kroz epohe bila je 32. Model je učen kroz 15 epoha gdje su praćene vrijednosti gubitka na skupu za učenje i na skupu za validaciju radi ranijeg zaustavljanja učenja.

Matrica zabune na skupu za testiranje nalazi se na Slika 4.3. Ovaj model postigao je točnost od 79 %, preciznost od 32,86 %, odziv od 23,71 % i F1 mjeru od 27,54 %. Po ovim rezultatima se vidi da i već vrlo jednostavan konvolucijski model može parirati rezultatima algoritama strojnog učenja na ovom problemu.



Slika 4.3 Matrica zabune jednodimenzionalnog konvolucijskog modela

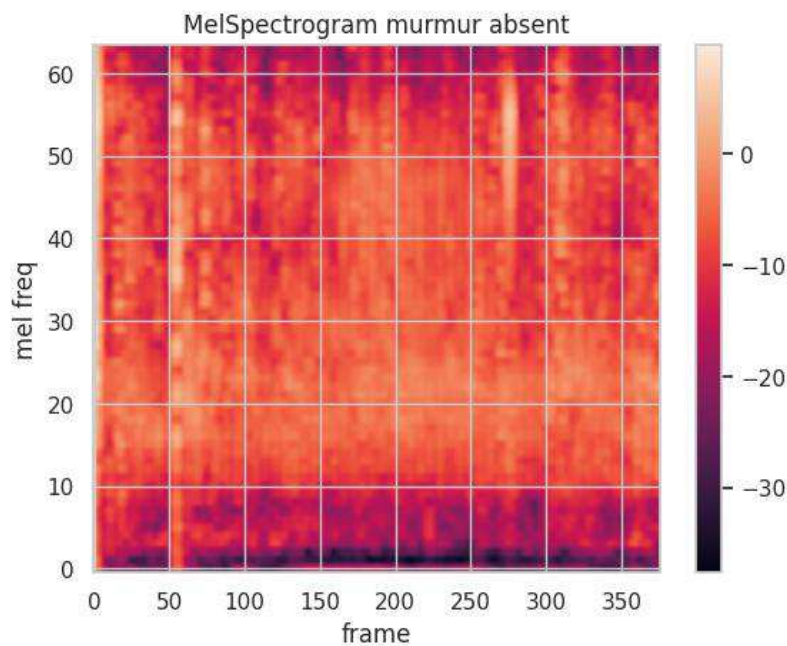
4.2.2. Dvodimenzionalni konvolucijski model

Dvodimenzionalnim konvolucijskim modelima na ulaz se ne može postaviti čisti audio zapis, ali audio zapisi se mogu prikazati na različite načine.

Način na koji se audio zapis može reprezentirati u dvije dimenzije je pomoću spektrograma. Spektrogram je, u praksi, jedan od najčešćih načina reprezentacije audio zapisa za njihovo korištenje u dubokom učenju. On prikazuje audio zapis istovremeno u frekvencijskoj i u vremenskoj domeni. Spektrogram je, također kao i kepralni koeficijenti mel-frekvencije, baziran na metodi FT opisanoj u poglavlju 3.1.1.

U ovom istraživanju korišteni su mel-spektrogrami. Prefiks mel dolazi od mel-skale koja je već objašnjena u poglavlju 3.1.2. Spektrogram nastaje provođenjem brze Fourierove transformacije kroz cijeli audio zapis s određenom manjom veličinom prozora koji se onda pomiče slijeva nadesno. Dobivene amplitude konvertiraju se u decibele (dB), a frekvencije se skaliraju po mel-skali. Tako dobiveni spektrogram naziva se mel-spektrogram i prikazuje zastupljenost različitih frekvencija u audio zapisu kroz vrijeme.

Za konverziju audio zapisa u mel-spektrogram potrebno je zadati nekoliko parametara: broj brzih Fourierovih transformacija, duljinu pomaka prozora, te broj mel-filtara. Za broj brzih Fourierovih transformacija korištena je uobičajena vrijednost od 512, duljina pomaka prozora je 32, a broj mel-filtara 64. Zapisi su prije izlučivanja spektrograma poduzorkovani na 1000 Hz, te su dovedeni na jednaku duljinu. Slika 4.4 prikazuje mel-spektrogram zapisa PCG u kojem nije prisutan srčani šum.

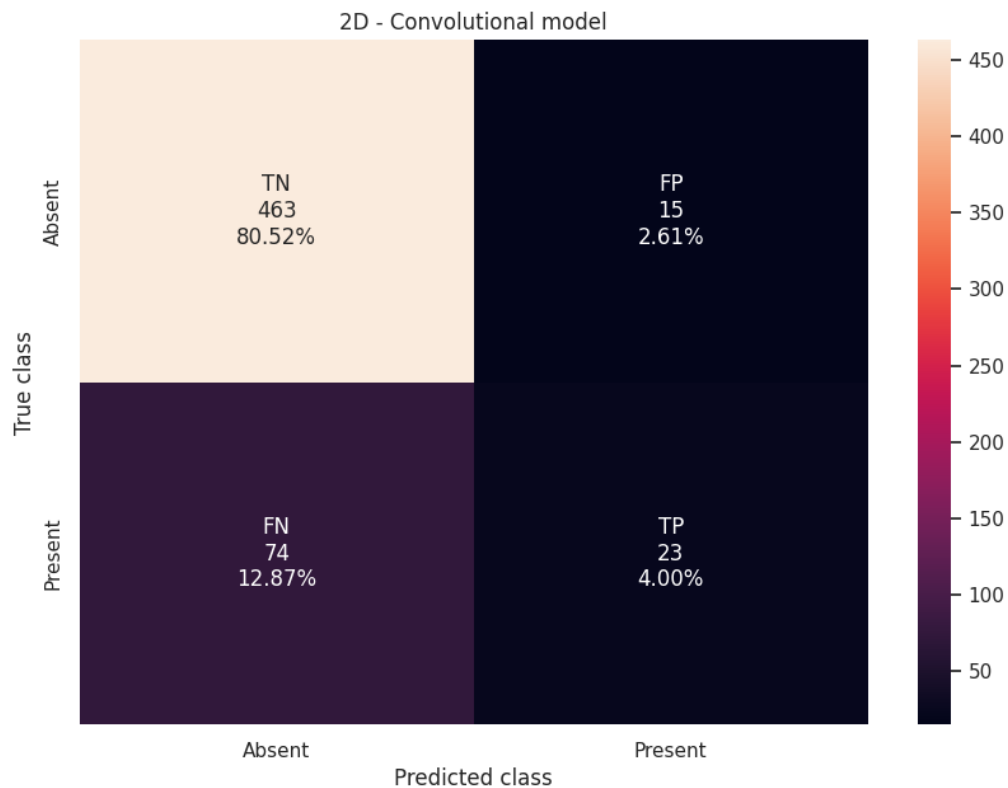


Slika 4.4 Mel-spektrogram PCG zapisa

Arhitektura dvodimenzionalnog konvolucijskog modela slična je onoj kod jednodimenzionalnog, samo što ovaj model koristi dvodimenzionalne konvolucije i ima četiri konvolucijska sloja umjesto dva koliko je imao prijašnji model. Hiperparametri za učenje modela ostaju isti.

Iz matrice zabune na Slika 4.5, vidi se da ovaj model na skupu za testiranje postiže točnost od 84,52 %, preciznost od 60,53 %, odziv od 23,71 % i F1 mjeru od 34,07 %. U usporedbi s prošlim modelom brojnost lažno pozitivnih primjera se jako smanjila, što je dovelo do

poboljšanja u rezultatu preciznosti za 27,67% i F1 mjere za 6,53%. Iz ovoga se zaključuje kako prelazak u frekvencijsku domenu omogućuje konvolucijskom modelu da izluči malo korisnije značajke.



Slika 4.5 Matrica zabune dvodimenzionalnog konvolucijskog modela

4.2.3. AlexNet

Osim klasične metode učenja modela, postoji i metoda učenja prijenosom (engl. *transfer learning*) u kojoj se koristi model koji je već naučen za rješavanje nekog srodnog problema. Pokretanje učenja takvog modela na novom skupu podataka rezultira finim podešavanjem parametara za rješavanje novog problema. Takvo učenje traje kraće i korisno je ako za novi problem postoji premalo podataka. Zbog nedostatka prednaučenih modela iz područja fonokardiografije, odabire se AlexNet kao kandidatni model za poboljšanje rezultata prijašnjih modela zbog njegove česte korištenosti u problemima kod klasifikacije slika.

AlexNet je konvolucijska neuronska mreža koju je predstavio Alex Krizhevsky 2012. godine i od tada je ova arhitektura poznata i često korištena za klasifikaciju i prepoznavanje slika (Krizhevsky et al., 2012). Naučena je na velikom skupu podataka ImageNet (Deng et al., 2009) koji sadrži 15 milijuna označenih slika visoke rezolucije koje pripadaju u otprilike

22.000 kategorija. AlexNet je prednaučen na podskupu ImageNeta koji sadrži otprilike 1,2 milijuna slika podijeljenih u 1000 kategorija (Krizhevsky et al., 2012).

Originalna mreža sadrži osam slojeva s težinama. Prvih pet su konvolucijski, a preostala tri su potpuno povezana. Izlaz posljednjeg potpuno povezanog sloja dovodi se do 1000-smjerne funkcije *softmax* koja proizvodi distribuciju preko 1000 oznaka klasa (Krizhevsky et al., 2012). Mreža koja je korištena za klasifikaciju PCG zapisa u ovom istraživanju morala je biti modificirana tako da na izlazu ima jedan neuron zbog zadatka detekcije srčanih šumova koji je binaran. Također funkciju *softmax* trebalo je zamijeniti sigmoidom. Na ulazu u mrežu potrebno je bilo prilagoditi prvi sloj da može primiti jednokanalne slike (spektrograme), a ne RGB slike koje sadrže 3 kanala. Arhitektura modificiranog AlexNeta za klasifikaciju zapisa PCG-a prikazana je na Slika 4.6.

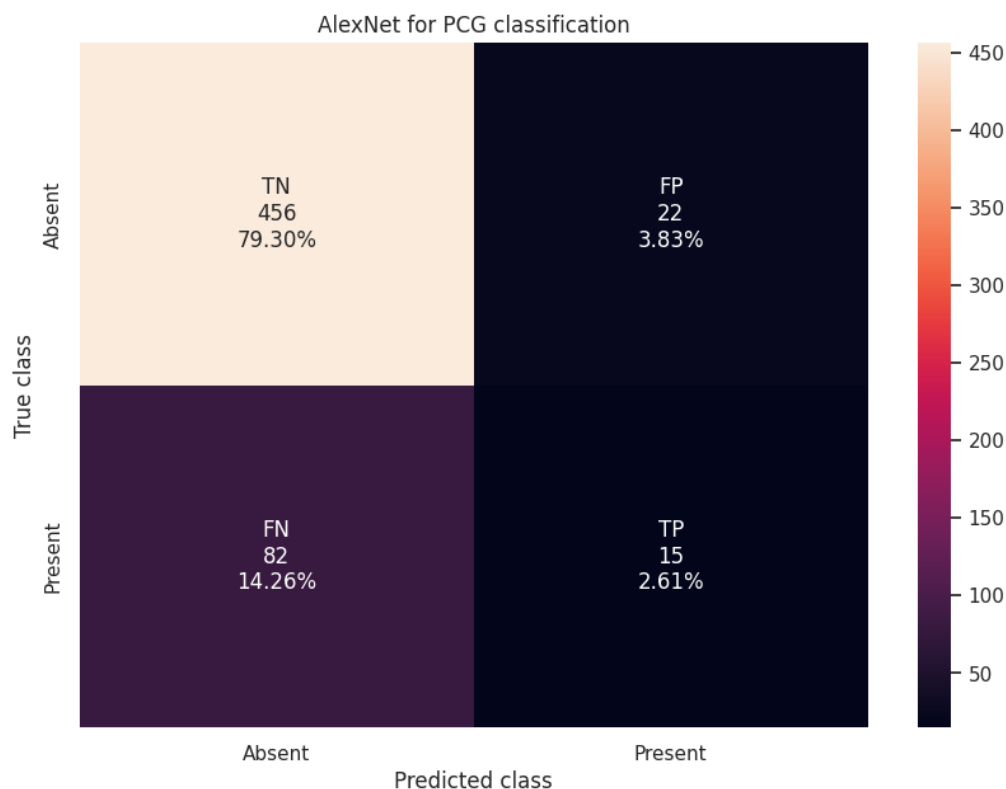
```
AlexNet(  
  (features): Sequential(  
    (0): Conv2d(1, 64, kernel_size=(11, 11), stride=(4, 4), padding=(2, 2))  
    (1): ReLU(inplace=True)  
    (2): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)  
    (3): Conv2d(64, 192, kernel_size=(5, 5), stride=(1, 1), padding=(2, 2))  
    (4): ReLU(inplace=True)  
    (5): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)  
    (6): Conv2d(192, 384, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))  
    (7): ReLU(inplace=True)  
    (8): Conv2d(384, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))  
    (9): ReLU(inplace=True)  
    (10): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))  
    (11): ReLU(inplace=True)  
    (12): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)  
  )  
  (avgpool): AdaptiveAvgPool2d(output_size=(6, 6))  
  (classifier): Sequential(  
    (0): Linear(in_features=9216, out_features=1, bias=True)  
    (1): Sigmoid()  
  )  
)
```

Slika 4.6 Modificirana arhitektura AlexNeta

AlexNetov model na svoj ulaz prima slike dimenzija 227x227. Mel-spektrogrami kao dvodimenzionalne reprezentacije zapisa PCG-a, sukladno tome, razvučeni su na tu rezoluciju. Ovaj postupak mijenja formalnu interpretaciju spektrograma, jer razvlači njegove osi, pa tako spektrogram više nije dobar za pravilno očitavanje zastupljenosti frekvencijskih komponenti signala. Unatoč tome, to neće utjecati na sposobnost konvolucijskog modela da iz takvog spektrograma izvuče korisne značajke, jer formalna interpretacija spektrograma konvolucijskom modelu ništa ne znači. On takav spektrogram promatra isključivo kao niz piksela.

Ovaj model optimiran je Adamom s istim parametrima kao i prijašnji modeli. Učen je kroz 40 epoha s veličinom grupe 32.

Na Sliku 4.7 prikazana je matrica zabune na skupu za testiranje.



Slika 4.7 Matrica zabune modificiranog AlexNet modela

Učenje je rezultiralo s točnosti od 81,91 %, preciznošću od 40,54 %, odzivom od 15,46 % i F1 mjerom od 22,39 %. Iako ima manje lažno pozitivno klasificiranih zapisa od jednodimenzionalnog konvolucijskog modela, rezultati pokazuju da se ovaj model nije dobro prilagodio zadatku detekcije srčanih šumova. Njegovi rezultati lošiji su od prijašnja dva konvolucijska modela.

4.2.4. ResNet

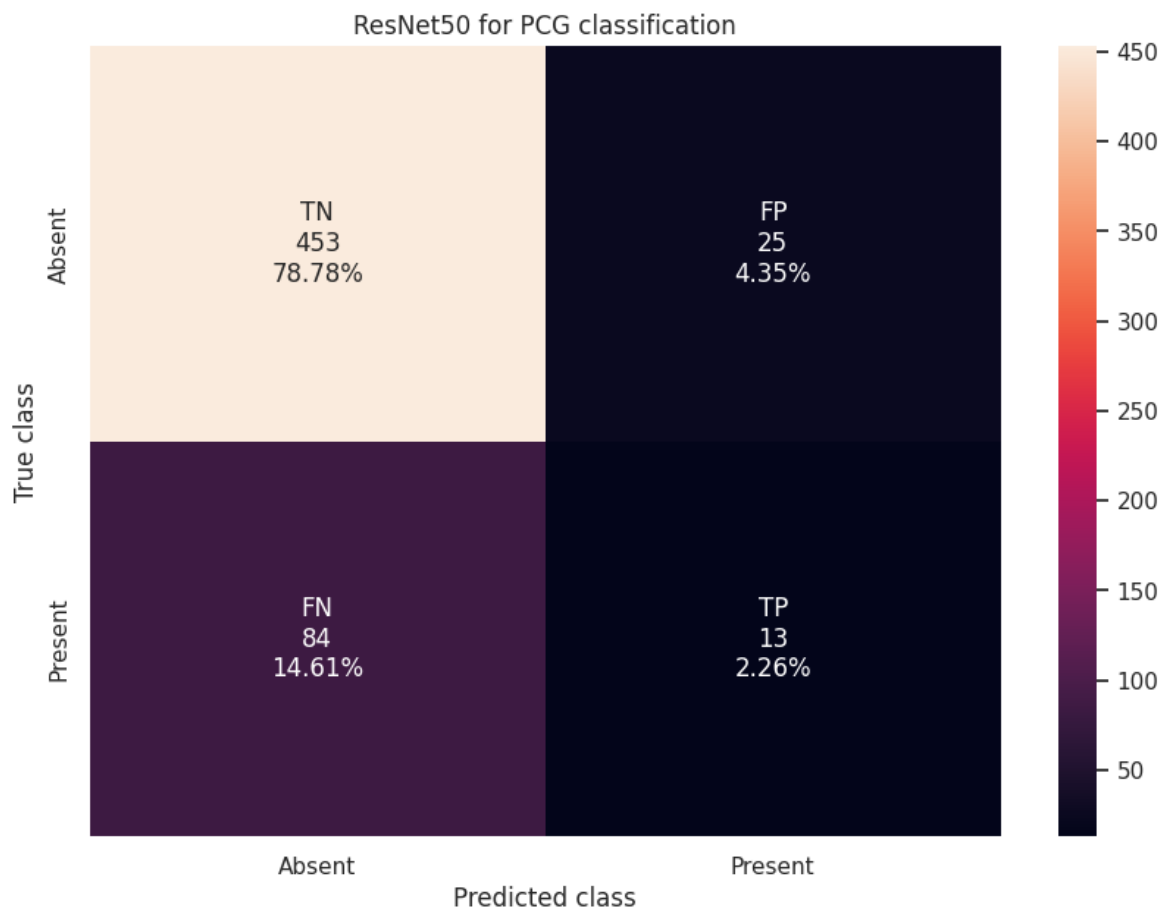
Residual Network (ResNet) je još jedan od „state-of-the-art“ dubokih konvolucijskih modela za zadatke klasifikacije slika. Ovakvu arhitekturu prvi put je predstavio Kaiming He 2016. godine (He et al., 2016). ResNet koristi arhitekturu rezidualnih veza s ciljem olakšavanja i signifikantnog ubrzanja učenja dubokih mreža. Ovakvim pristupom omogućeno je učenje mreža koje su znatno dublje od prethodno korištenih modela. Slojevi u ovoj arhitekturi formulirani su kao rezidualne funkcije (He et al., 2016). Općenita rezidualna funkcija prikazana je formulom (4.1).

$$y = x + f(x) \quad (4.1)$$

Formulacija (4.1) može se realizirati neuronskim mrežama s "prečacima". Prečaci su veze koje preskaču jedan ili više slojeva u mreži. U slučaju ResNeta, takve veze jednostavno izvode funkciju identiteta, a njihovi se izlazi dodaju na izlaze sekvence konvolucijskih slojeva. Drugim riječima, na izlaze konvolucijske sekvence, dodaje se ulaz u tu sekvencu (He et al, 2016).

Postoje mnoge verzije ResNet arhitekture. U ovom istraživanju ispitana je modificirana ResNet50 arhitektura. Naziva se ResNet50 jer sadrži 50 slojeva, od kojih su 48 konvolucijski slojevi, jedan sloj sažimanja maksimalnom vrijednošću i jedan sloj sažimanja srednjom vrijednošću. Modificirana je tako da na ulaz prima sive jednokanalne slike umjesto slika u boji s 3 kanala. Također je izmijenjen izlaz mreže da klasificira spektrograme u dvije klase. Spektrogrami su razvučeni na veličinu 224x224 piksela, kako bi odgovarali arhitekturi mreže.

Ovaj model učen je kroz 40 epoha, veličinom grupe od 32 primjera i optimizatorom Adam. Postigao je točnost od 81,04 %, preciznost od 34,21 %, odziv od 13,40 % i F1 mjeru od 19,26 % na skupu za testiranje. Na Slika 4.8 prikazana je matrica zabune.



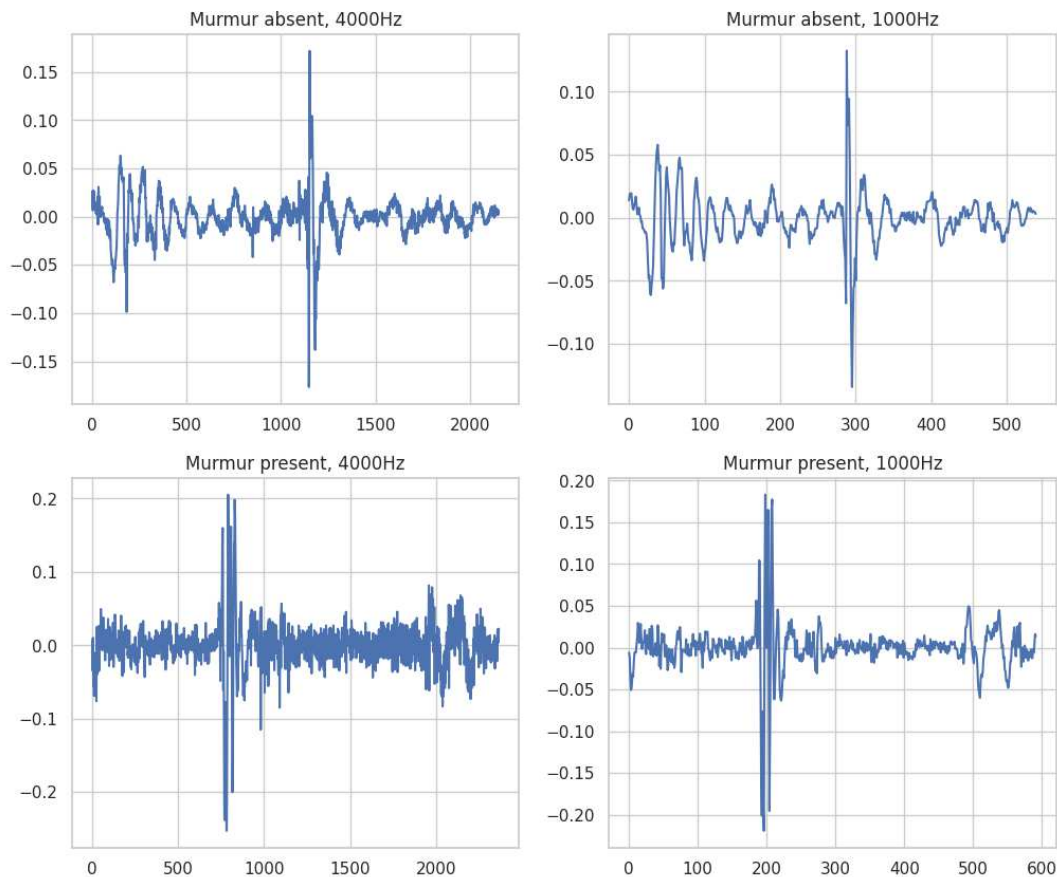
Slika 4.8 Matrica zabune modificiranog ResNet50 modela

Po rezultatima klasifikacije moguće je dati isti zaključak kao i za model AlexNet. Ovim modelom nije postignut napredak u detekciji srčanih šumova.

4.2.5. Metoda podjele PCG-a na srčane cikluse

Srčani ciklus definira se kao vremenski period od početka jednog otkucaja srca do početka sljedećeg. Svaki srčani ciklus započinje S1 zvukom, nakon kojeg slijedi sistolički period. Sistolički period završava se S2 zvukom, nakon kojeg slijedi dijastolički period kao završni segment u ciklusu. U poglavlju 2.4 opisano je kako svaki zapis PCG-a dolazi s odgovarajućom segmentacijskom datotekom formata naziva `ABCDE_XY.tsv`. Ta datoteka sadrži informacije s točnim vremenima pojavljivanja srčanog zvuka S1, sistole, srčanog zvuka S2 i dijastole. Datoteke su generirane posebnim algoritmima za detekciju ovih događaja u PCG zapisu. To omogućuje podjelu svakog zapisa na više srčanih ciklusa, ovisno o tome koliko ih sam zapis sadrži. Slika 4.9 prikazuje srčane cikluse bez i sa srčanim šumom, te njihove poduzorkovane varijante koje se koriste za učenje modela.

Cardiac cycles



Slika 4.9 Primjeri srčanih ciklusa

Analizom segmentacijskih datoteka uočeno je kako ponekad u zapisima postoje nedefinirani vremenski intervali koje algoritam nije prepoznao kao jednog od dijelova srčanog ciklusa. Također postoje ciklusi koji ne sadrže sve dijelove koje bi srčani ciklus morao imati (na primjer, nakon sistoličnog perioda dolazi S1 zvuk, umjesto zvuka S2). Pretpostavka je da se to događa zbog podatkovnog šuma koji je prisutan u skupu podataka. Moguće je da je kvaliteta nekog segmenta narušena zbog raznih pozadinskih zvukova i faktora okoline u kojoj je zapis snimljen.

Sukladno ovome, predstavlja se metoda detekcije srčanih šumova podjelom zapisa PCG-a na srčane cikluse. Svaki zapis podijeljen je na srčane cikluse prema segmentacijskoj datoteci na način da su uzeti samo oni ciklusi koji sadrže sve četiri nužne komponente: S1, sistolu, S2 i dijastolu. Ovim uvjetom se postiže da se iz podataka uklanjaju neprepoznatljivi segmenti zapisa, te ciklusi s neprepoznatljivim komponentama.

Ovom metodom mreže su učene na srčanim ciklusima, a ne na cijelom zapisu PCG-a odjednom. Svakom ciklusu pridjeljuje se predikcija modela, a na kraju u fazi testiranja

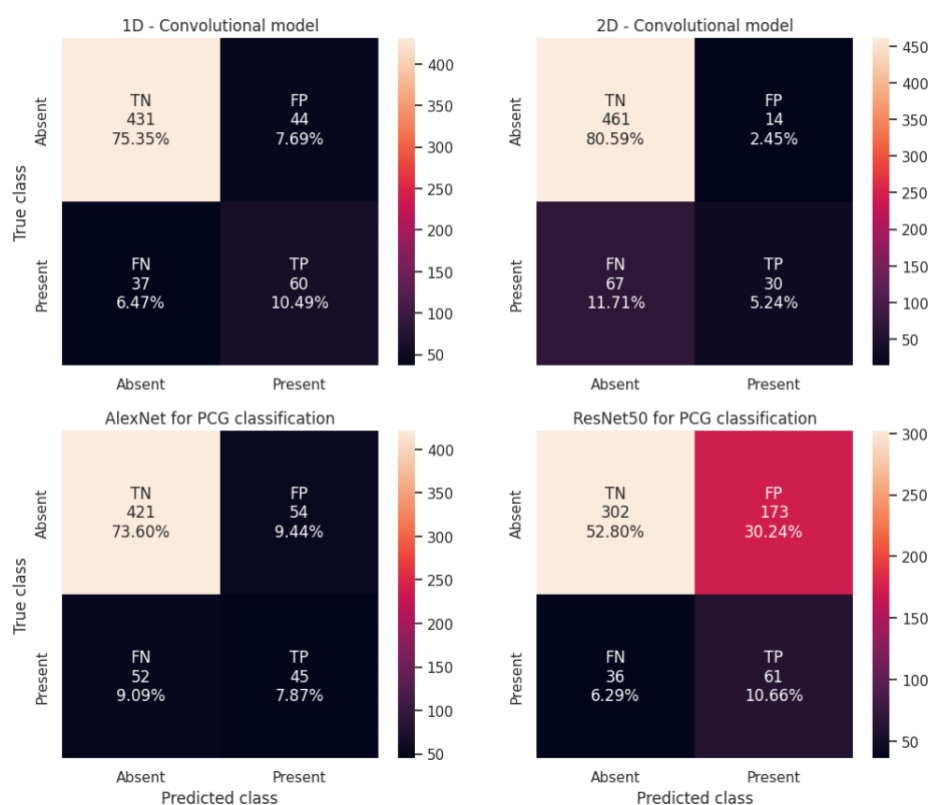
modela uzimaju se u obzir predikcije svih ciklusa iz određenog zapisa PCG-a, te se glasanjem odlučuje konačna predikcija za taj zapis. Ako je za više od 40% srčanih ciklusa u zapisu detektiran srčani šum, onda se zapis klasificira pozitivno, u suprotnom, klasificira se u negativnu klasu. Ako je u pojedinim ciklusima detektiran srčani šum, a u nekima nije, mala prednost se daje pozitivnoj klasifikaciji. Zbog toga je granica odlučivanja 40%. Ovime se odziv modela trebao poboljšati, što je važno za modele koji se bave medicinskim dijagnozama.

Pretpostavka je da bi ovakva metoda detekcije srčanih šumova mogla smanjiti pristranost modela većinskoj klasi, poboljšati varijabilnost, kao i njegov odziv i konačni rezultat, zbog uklanjanja određenog šuma u podacima i načina donošenja odluke o ciljnoj varijabli.

Metoda je ispitana na prijašnjim konvolucijskim modelima, s time da jednodimenzionalni model uči na jednodimenzionalnim podacima zvuka, dok dvodimenzionalan model, AlexNet i ResNet50 uče na spektrogramima.

Slika 4.10 prikazuje četiri matrice zabune iz kojih se vide rezultati ove metode na prijašnjim konvolucijskim modelima. Usporedbe radi, u Tablica 2 dani su rezultati ovih modela klasičnom metodom učenja te metodom podjele podataka na srčane cikluse.

Cardiac cycle segmentation method



Slika 4.10 Matrice zabune za konvolucijekse modele metodom podjele na srčane cikluse

Tablica 2 Rezultati konvolucijskih modela

| <i>Neuronska mreža</i> | Točnost | Preciznost | Odziv | F1 mjera |
|--------------------------------|----------------|-------------------|--------------|-----------------|
| <i>1D konvolucijski model</i> | 78,96 % | 32,86 % | 23,71 % | 27,54 % |
| <i>1D konvolucijski model*</i> | 85,84 % | 57,69 % | 61,86 % | 59,70 % |
| <i>2D konvolucijski model</i> | 84,52 % | 60,53 % | 23,71 % | 34,07 % |
| <i>2D konvolucijski model*</i> | 85,84 % | 68,18 % | 30,93% | 42,55 % |
| <i>Modificirani AlexNet</i> | 81,91 % | 40,54 % | 15,46 % | 22,39 % |
| <i>Modificirani AlexNet*</i> | 81,49 % | 45,45 % | 46,39 % | 45,92 % |
| <i>Modificirani ResNet50</i> | 81,04 % | 34,21 % | 13,40 % | 19,26 % |
| <i>Modificirani ResNet50*</i> | 63,46 % | 26,07 % | 62,89 % | 36,86 % |

* označeni modeli učeni su metodom podjele na srčane cikluse

Prateći F1 mjeru, zaključuje se kako su mreže učene metodom podjele na srčane cikluse dale bolje rezultate od mreža koje su učile na kompletnim zapisima. Ova metoda dala je bolje rezultate, poboljšanjem odziva i F1 mjere kod svih modela. Preciznost se također povećala kod svih modela osim ResNeta, koji je imao povećani broj lažno pozitivnih primjeraka, što je utjecalo na smanjenje točnosti kod ovog modela.

Najveći napredak ostvaren je na jednodimenzionalnoj konvolucijskoj mreži čiji su se odziv i F1 mjera povećali za čak 38,15 % i 32,16 %. Ovo je trenutno najuspješniji model s točnošću 85,84 % i F1 mjerom 59,70 %.

4.3. Povratni modeli

Kao i konvolucijski modeli, povratni modeli spadaju u kategoriju algoritama dubokog učenja. Povratni modeli omogućuju obradu podataka sekvencijske strukture. Takvi podaci imaju neku vrstu vremenske komponente. To na primjer može biti broj riječi u rečenici ili vremenski niz koji je slijed određenih vrijednosti kroz vrijeme. Tako su audio signali poput zapisa PCG-a pogodni za obradu ovakvim modelima.

Povratni modeli sastoje se od slojeva koji sadrže povratne ili rekurentne ćelije. Kao i potpuno povezani sloj neuronske mreže, svi neuroni rekurentne ćelije imaju i dolazne veze koje proizlaze iz svih neurona prethodnog sloja i odlazne veze koje vode do svih neurona u sljedećem sloju. Za razliku od potpuno povezanih slojeva, rekurentni slojevi također imaju veze između neurona unutar istog sloja, koje se nazivaju povratnim vezama. Potpuno povezana rekurentna ćelija u sebi ima međusobne veze od svakog neurona do svakog drugog neurona (uključujući veze neurona sa samim sobom). Stoga rekurentni sloj s r neurona ima ukupno r^2 rekurentnih veza (Buduma et al., 2022).

Budući da su podaci koje povratne ćelije obrađuju sekvence, proces obrade može se podijeliti u vremenske korake. U svakom vremenskom koraku, model na ulaz dobiva sljedeći element sekvence. Dolazne veze iz neurona prethodnog sloja komponenta su ćelije koja prenosi informacije izračunate iz trenutnog vremenskog koraka. Povratne veze, međutim, prenose informacije pohranjene iz prethodnog vremenskog koraka (Buduma et al., 2022). To znači da je izlaz iz povratnog sloja zapravo funkcija njegovih ulaza i izlaza iz prošlog vremenskog koraka. Na ovaj način model preko povratnih veza čuva takozvano skriveno stanje gdje pohranjuje i pamti informacije o dosad viđenim podacima. Zbog toga ovakvi modeli mogu modelirati vremenske ovisnosti u podacima.

Formula (4.2) prikazuje postupak izračuna skrivenog stanja (h^t) jednostavne povratne ćelije u vremenskom trenutku t . W_{hh} označava veze između neurona u istom sloju, dok W_{hx} označava veze neurona trenutnog i prošlog sloja, a b_h je vektor pristranosti. g je neka nelinearna aktivacijska funkcija. Kod povratnih mreža je to najčešće tangens hiperbolni ili sigmoida.

$$h^{(t)} = g(W_{hh}h^{(t-1)} + W_{hx}x^{(t)} + b_h) \quad (4.2)$$

Moguće je povezati više povratnih slojeva jedan za drugim, što se naziva višeslojnim povratnim modelom. Ako se ovakva mreža koristi za problem klasifikacije, na kraju se dodaje potpuno povezani sloj koji obavlja klasifikaciju prema zadnjem skrivenom stanju.

Prvi ispitani povratni model za detekciju srčanih šumova je višeslojni model koji koristi jednostavne povratne ćelije. Ovaj model na svom ulazu prima audio zapise u jednodimenzijском obliku vremenskog niza, isto kao i jednodimenzionalan konvolucijski model. Taj model ima 4 povratna sloja s veličinom skrivenog stanja 600, te potpuno povezani sloj s 128 neurona i zadnji potpuno povezani sloj s jednim neuronom za klasifikaciju. Optimiran je algoritmom Adam, učen kroz 15 epoha s veličinom grupe 32.

Model postiže točnost na skupu za testiranje od 78,26 %, preciznost od 26,67 %, odziv od 16,49 % i F1 mjeru 20,38 %. Na Slika 4.11 je prikazana matrica zabune.

Ako se u modelu koriste klasični povratni slojevi opisani u ovom poglavlju, moguće je da će se dogoditi problem nestajućeg ili eksplodirajućeg gradijenta. Ovaj problem javlja se u fazi prolaza mrežom unatrag, kada se računaju gradijenti, a nastaje zbog povratnih veza i uzastopnog množenja matricom W_{hh} . Ovo znatno otežava učenje modela zasnovanih na ovakvim povratnim ćelijama. Motivirani ovime, znanstvenici su predstavili drugačije ćelije koje umanjuju ovaj problem.

4.3.1. Ćelija s dugom kratkoročnom memorijom

Ćelija s dugom kratkoročnom memorijom (engl. *long short-term memory*, LSTM) spada u napredne povratne ćelije, uvedena je kao rješenje problema nestajućeg gradijenta. Predstavili su ju Hochreiter i Schmidhuber 1997. godine (Hochreiter & Schmidhuber, 1997). Osnovna ideja LSTM-arhitekture bila je da mreža može pouzdano prenositi važne informacije mnogo vremenskih koraka u budućnost (Buduma et al., 2022).

Slično kao skriveno stanje kod jednostavne povratne ćelije, takozvana memorijska ćelija (engl. *memory cell*) sastavni je dio LSTM arhitekture koji čuva kritične informacije naučene tijekom vremena. Mreža je oblikovana da na učinkovit način održava korisne informacije u memorijskoj ćeliji tijekom mnogih vremenskih koraka. Novosti kod LSTM-a s obzirom na jednostavnu povratnu ćeliju su vrata ili propusnice koje na određen način filtriraju podatke. Propusnica zaboravljanja (engl. *keep gate*) odlučuje koje informacije iz memorijske ćelije će zadržati, a koje zaboraviti. Propusnica novog ulaza (engl. *write gate, input gate*) odlučuje koje informacije će propustiti u memorijsku ćeliju a koje će odbaciti. Izlazna propusnica (engl. *output gate*) daje izlaz iz LSTM sloja, koji označava interpretaciju vektora stanja (Buduma et al., 2022).

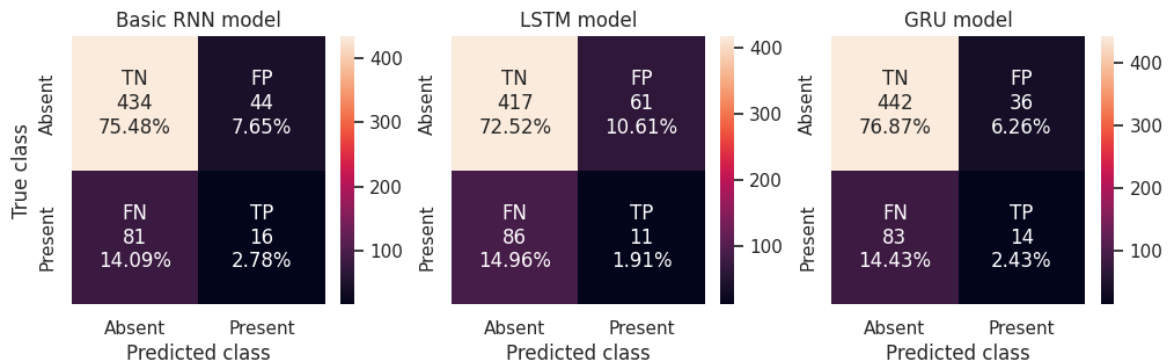
Modeli LSTM-a poznati su po dužem vremenu učenja pa je arhitektura ispitnog LSTM modela smanjena s obzirom na jednostavni povratni model. Ispitani model ima dva LSTM-sloja s veličinom skrivenog stanja 200 jedinica, potpuno povezani sloj sa 128 neurona i zadnji sloj za binarnu klasifikaciju s jednim neuronom. Model je učen kroz 30 epoha s veličinom grupe 32. Optimiran je Adamom, a na ulazu prima čiste jednodimenzionalne signale PCG-a. Mreža na skupu za testiranje postiže lošije rezultate od jednostavnog povratnog modela. Smanjena arhitektura ove mreže mogući je razlog i manje točnosti u odnosu na jednostavan povratni model, jer takva mreža ima manji kapacitet. Vrijednost točnosti je 74,43 %, preciznost 15,28 %, odziv 11,34 %, a F1 mjera 13,02 %. Slika 4.11 prikazuje matricu zabune LSTM modela na skupu za testiranje.

4.3.2. Propusna povratna ćelija

Propusna povratna ćelija (engl. *gated recurrent unit, GRU*) novija je varijanta napredne povratne ćelije. GRU arhitekturu 2014. razvio je Kyunghyun Cho (Cho et al., 2014). GRU također koristi propusnice koje filtriraju podatke u mreži. Sastoji se od propusnice ažuriranja (engl. *update gate*) i propusnice resetiranja (engl. *reset gate*). Prije izračuna skrivenog stanja, ova ćelija koristi privremeno stanje kao međukorak u izračunu.

U praksi se koristi jednako kao i LSTM, ovisno o preferencijama, ali sadrži manje parametara zbog manje propusnica i vektora, pa je proces učenja povratne mreže GRU brži. GRU-ov model jednake je arhitekture kao i model s jednostavnim povratnim ćelijama. Ima četiri povratna sloja s ćelijama GRU, jedan potpuno povezani sloj sa 128 neurona i zadnji sloj s jednim neuronom koji klasificira dobivene podatke preko sigmoidne funkcije. Učen je

kroz 12 epoha i veličinom grupe 32, optimizatorom Adam. Na skupu za testiranje postiže točnost od 79,30 %, preciznost od 28 %, odziv od 14,43 % i F1 mjeru od 19,05 %. Matrica zabune ovog modela prikazana je Slika 4.11.



Slika 4.11 Matrice zabune povratnih modela

Od tri ispitane povratne neuronske mreže, najbolje rezultate ima model s jednostavnim povratnim ćelijama, iako su ti rezultati vrlo slični. Rezultati kod povratnih modela lošiji su od rezultata konvolucijskih modela pa bi se dalo zaključiti kako su konvolucijski modeli prikladniji za rješavanje problema detekcije srčanih šumova.

4.4. Kombinacija konvolucijskih i povratnih slojeva

Iako su konvolucijske mreže bolje klasificirale PCG zapise od povratnih mreža, one se u praksi često koriste zajedno u istoj mreži. Na početku mreže nalaze se konvolucijski slojevi, čiji se izlazi propagiraju kroz povratne slojeve i na kraju klasificiraju pomoću izlaznih potpuno povezanih slojeva.

Konvolucijski slojevi pridonose izlučivanjem bitnih značajki, dok povratni slojevi kroz vrijeme čuvaju informacije. Takve kombinacije u praksi često donose dobre rezultate u raznim primjenama.

Ispitani model spoj je dva najbolja modela od konvolucijskih i povratnih arhitektura. Na ulazu koristi dva sloja jednodimenzionalne konvolucije sa veličinom jezgre 5, pomakom 1 i nadopunjavanjem 2. Svaki konvolucijski sloj ima sloj normalizacije po grupi i sloj sažimanja maksimalnom vrijednošću. Nakon konvolucije dolaze dva sloja s jednostavnim povratnim ćelijama s veličinom skrivenog stanja 400. Na kraju su dva potpuno povezana sloja s 512 neurona i 1 neuronom. Arhitektura ovog modela dana je Slika 4.12.

```

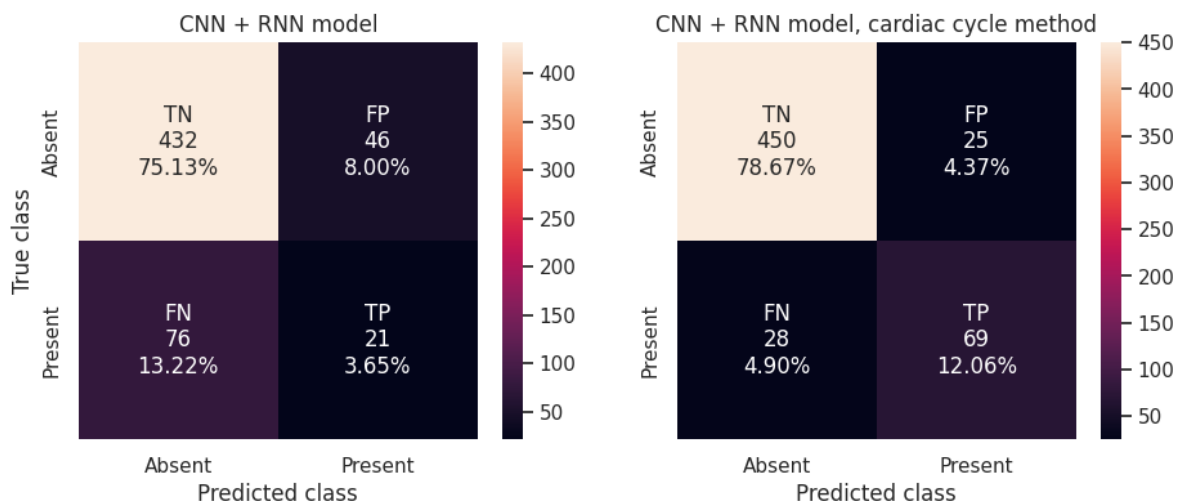
CardiacCycleCNN_RNNmodel(
  (conv1): Conv1d(1, 16, kernel_size=(5,), stride=(1,), padding=(2,))
  (bn1): BatchNorm1d(16, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (maxpool1): MaxPool1d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (conv2): Conv1d(16, 32, kernel_size=(5,), stride=(1,), padding=(2,))
  (bn2): BatchNorm1d(32, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (maxpool2): MaxPool1d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (rnn1): RNN(6560, 400, batch_first=True)
  (rnn2): RNN(400, 400, batch_first=True)
  (fc1): Linear(in_features=400, out_features=512, bias=True)
  (fc2): Linear(in_features=512, out_features=1, bias=True)
)

```

Slika 4.12 Arhitektura modela s konvolucijskim i povratnim slojevima

Model je prvo ispitan klasičnom metodom učenja s cijelim zapisima PCG-a, a nakon toga učen je metodom podjele zapisa PCG-a na srčane cikluse opisane u poglavlju 4.2.5.

Učenjem klasičnom metodom, na skupu za testiranje ovaj model postigao je točnost od 78,78 %, preciznost od 31,34 %, odziv od 21,65 % i F1 mjeru od 25,61 %. Učenjem na srčanim ciklusima model je ostvario točnost od 90,73 %, preciznost od 73,40 %, odziv od 71,13 % i F1 mjeru od 72,25 %. Na slici su prikazane matrice zabune za ova dva modela.



Slika 4.13 Matrice zabune za konvolucijsko-rekurentne modele

Model koji je učio na srčanim ciklusima ima znatno veću uspješnost klasifikacije. Ovakav napredak u rezultatu pokazuje korisnost kombinacije konvolucijskih i rekurentnih slojeva. Mreža je uspjela naučiti obrasce kojima vrlo uspješno klasificira zapise PCG-a. Za sada su ovo najbolji rezultati svih metrika.

5. Transformerski modeli

Izumom transformerske arhitekture, područje umjetne inteligencije jako je napredovalo. Transformeri su se pokazali moćnim i preciznim alatom u mnogim sferama kojima se umjetna inteligencija bavi. Glavno područje primjene transformera je obrada prirodnog jezika (engl. *natural language processing*), ali s vremenom su razvijene različite modifikacije ovih arhitektura kako bi se njihov doseg proširio i na druga područja. Tako su transformeri danas uspješni i u računalnom vidu, sustavima za preporučivanje, financijama, ali najvažnije za ovo istraživanje, u obradi zvuka.

Prva transformerska arhitektura predstavljena je u članku Ashisha Vaswanija 2017. godine s pojmom pozornosti kao najvažnijim mehanizmom u transformerskoj arhitekturi (Vaswani et al., 2017). Ključni zaključak ovog članka je da se pozornost može koristiti kao jedini mehanizam za opisivanje ovisnosti između podataka na ulazu i izlaza modela, što znači eliminaciju povratnih slojeva poput LSTM-a ili GRU-a (Amatriain, 2023). Naime, za razliku od povratnih modela koji obrađuju podatke sekvencijalno od početka do kraja, transformeri obrađuju cijelu sekvencu odjednom. Uz pomoć mehanizma pozornosti, pamte odnose između bilo kojih elemenata u ulaznoj sekvenci, dok povratni modeli imaju ograničen doseg pamćenja informacija.

Transformerska arhitektura sastoji se od mreže čitača ili kodera (engl. *encoder*) i mreže pisara ili dekodera (engl. *decoder*). Uloga kodera je da dobivenu ulaznu sekvencu pretvori u sažetu reprezentaciju koju će dekodek koristiti za generiranje izlazne sekvence. Ulaz u transformer je sekvenca pretvorena u niz tokena operacijom vektorske reprezentacije (engl. *embedding*). Izlaz koderu je prikaz fiksne dimenzije za svaki od tokena, s odvojenom vektorskom reprezentacijom za niz kao cjelinu. Dekoder uzima izlaz koderu kao ulaz i izbacuje niz tokena kao svoj izlaz. Tokeni su često riječi ili podriječi ako se radi o obradi prirodnog jezika (Amatriain, 2023).

U originalnom članku koder i dekodek imali su šest jednakih slojeva. Koder je u svakom od tih slojeva imao sloj samo-pozornosti i potpuno povezani sloj. Sloj samo-pozornosti računa reprezentacije za svaki token na temelju svih ostalih ulaznih tokena. Nakon svakog podsloja dolazi rezidualni sloj koji olakšava računanje gradijenata i sloj normalizacije koji olakšava i ubrzava proces učenja modela. Arhitektura dekodera je slična koderu, ali on koristi dodatan sloj samo-pozornosti koji se razlikuje od onog u koderu. Mnogi detalji u implementaciji

transformera s vremenom su modificirani, pa tako postoje modeli poput BERT-a koji je temeljen samo na koderu ili poput GPT-a koji je temeljen samo na dekoderu (Amatriain, 2023).

Kako bi transformer mogao obrađivati zvuk, na njegov ulaz potrebno je dovesti vektorsku reprezentaciju sekvence. Takvi vektori dobivaju se korištenjem konvolucijskih slojeva koji izlučuju značajke iz audio signala i stvaraju vektore kao reprezentacije zvuka. Često se konvolucijski slojevi primjenjuju na spektrograme jer su oni vrlo informativna reprezentacija audio signala koja se može tretirati kao slika, a konvolucijski slojevi dobro rade sa slikama.

Danas postoje mnoge transformerske arhitekture koje su unaprijed naučene na velikim označenim i neoznačenim skupovima podataka. Takvi modeli se mogu lako podesiti dodatnim učenjem na novom skupu podataka u svrhu raznih primjena. Tako će u ovom istraživanju biti ispitani neki unaprijed naučeni modeli, finim podešavanjem za detekciju srčanih šumova iz snimki PCG-a. Budući da nema mnogo kvalitetnih modela unaprijed naučenih na PCG-snimkama, koristit će se modeli koji dobro rade s govorom, u nadi da će se dobro prilagoditi problemu detekcije srčanih šumova.

5.1. Wav2vec 2.0

Wav2vec 2.0 (skraćeno wav2vec2) implementirali su znanstvenici iz Facebooka 2020. godine (Baevski et al., 2020). Wav2vec2 transformerski je model koji koristi samo koderski dio arhitekture originalnog transformera. Karakterizira ga princip samonadziranog učenja. To je učenje općih prikaza podataka iz neoznačenih podataka i kasnije fino podešavanje modela na označenim podacima.

Wav2vec2 na početku koristi višeslojni konvolucijski koder značajki koji na ulazu uzima neobrađeni zvuk i daje latentne govorne reprezentacije za T vremenskih koraka. Koder se sastoji od nekoliko konvolucijskih slojeva praćenih slojem normalizacije i aktivacijskom funkcijom GELU. Početni audio zapis prije slanja u koder normiran je na nultu srednju vrijednost i jediničnu varijancu. Ukupni pomak konvolucijskog koderu određuje broj vremenskih koraka T , točnije duljinu sekvence koja ulazi u transformer. Transformer iz ulazne sekvence gradi reprezentacije zvuka kroz skrivena stanja koja može koristiti za transkripciju govora (Baevski et al., 2020).

Za samonadzirano učenje modela maskira se određeni udio vremenskih koraka u prostoru latentnih značajki kodera slično modeliranju maskiranog jezika kod BERT-a. Cilj modela je prepoznati točnu kvantiziranu latentnu audio reprezentaciju iz skupa distraktora za svaki maskirani vremenski korak. Nakon toga računa se takozvani kontrastni gubitak pomoću kojeg model uči (Baevski et al., 2020).

Ispitani model wav2vec2 unaprijed je naučen na audio zapisima iz skupa podataka LibriSpeech koji se sastoji od brojnih audio snimki izvedenih iz audio knjiga (Panayotov et al., 2015). Budući da su audio snimke iz tog skupa podataka uzorkovane frekvencijom od 16 kHz, zbog toga će i zapisi PCG-a morati biti naduzorkovani na 16 kHz. Budući da se ovaj model već ponaša kao klasifikator, jer uči prepoznavati maskirane dijelove zvuka, nisu potrebne velike preinake u arhitekturi da bi se koristio za detekciju srčanih šumova. Jedino što je potrebno promijeniti je broj izlaza zadnjeg sloja modela da odgovara klasificiranju u dvije klase, te korištenje uobičajene funkcije gubitka binarne unakrsne entropije za fino podešavanje modela, umjesto kontrastnog gubitka koje se koristi za samonadzirano učenje.

Ispitani model wav2vec2 učen je kroz 30 epoha algoritmom Adam sa stopom učenja $3e-5$. Na skupu za testiranje postigao je točnost od 78,96 %, preciznost od 42,94 %, odziv od 75,26 % i F1 mjeru od 54,68 %. Na Slika 5.1 prikazana je matrica zabune ovog modela na skupu za testiranje. Vidi se da model nije pristran prema većinskoj klasi, pa je dosta primjeraka klasificirao pozitivno. Uspio je detektirati srčani šum kod 75,26 % primjeraka koji ga zaista imaju, ali je unatoč tome lažno klasificirao 20,29 % negativnih primjeraka kao pozitivne. Ovaj model za sada postiže najbolji odziv od svih ispitanih, ali F1 mjera mu je lošija od konvolucijsko-povratnog modela učenog na srčanim ciklusima.

5.2. HuBERT

Hidden-Unit BERT, skraćeno HuBERT, model je koji je jako sličan modelu wav2vec2 iz prošlog poglavlja. Prestavljen je 2021. godine (Hsu et al., 2021). Arhitektura modela jednaka je onoj kod modela wav2vec2. HuBERT također koristi samo koderski dio transformera, u kojeg se šalju ugrađene reprezentacije zvuka dobivene konvolucijom iz čistog audio zapisa. Razlika kod ovih modela je u načinu samonadziranog učenja.

HuBERT-ov pristup samonadziranom učenju koristi algoritam grupiranja (engl. *clustering*) za generiranje usklađenih ciljnih oznaka korištene za gubitak predviđanja sličan BERT-u. Ključan korak ovog pristupa je primjena gubitka predviđanja samo na maskirana područja,

što rezultira time da model nauči kombinirani akustički i jezični model preko kontinuiranih ulaza. HuBERT se prvenstveno oslanja na dosljednost koraka nenadziranog grupiranja, a ne na intrinzičnu kvalitetu dodijeljenih oznaka klastera (Hsu et al., 2021).

Arhitektura ovog modela sastoji se od višeslojnog konvolucijskog kodera značajki kao i kod modela wav2vec2, nakon kojeg dolazi koder BERT. Izlazi iz koder BERT šalju se u projekcijski sloj i u sloj za ugrađivanje (Hsu et al, 2021). Na izlazu iz modela dobivaju se skrivena stanja koja su analogna tokenima u rečenici teksta.

U istraživanju je korišten model unaprijed naučen na 960 sati zvuka iz skupa podataka LibriSpeech. Njegov izlazni sloj promijenjen je kako bi služio za klasifikaciju u dvije klase. Model je učio kroz 30 epoha, sa stopom učenja $3e-5$ i optimizatorom Adam. Na skupu za testiranje postiže točnost od 88,70 %, preciznost od 63,56 %, odziv od 77,32 % i F1 mjeru od 69,77 % i tako pokazuje bolje rezultate od modela wav2vec2. Ovaj model trenutno ima najbolji odziv sa samo 22 lažno negativna primjerka od ukupno 97 pozitivnih u skupu za testiranje. Po F1 mjeri je još uvijek bolji konvolucijsko-povratni model učen na srčanim ciklusima i to za 2,48%. Na Slika 5.1 je prikazana matrica zabune ovog modela na skupu za testiranje.

5.3. Audio Spectrogram Transformer

Prijašnja dva modela primjenjivala su operaciju konvolucije na jednodimenzionalne nizove koji su predstavljali čisti audio zapis. Model *Audio Spectrogram Transformer* (AST) umjesto čistog zvuka radi sa spektrogramima. Glavna ideja ovog modela je rješavanje raznih problema obrade zvuka, bez konvolucijskih slojeva ispred transformerske arhitekture. Ovaj model predstavio je Yuan Gong 2021. godine (Gong et al., 2021).

Potaknuti dobrim rezultatima transformera u području računalnog vida, Yuan Gong i suradnici razvili su model AST po uzoru na arhitekturu Vision Transformera (ViT) (Dosovitskiy et al., 2020), koji također eliminira prijašnje korištene konvolucijske slojeve kod obrade slike, te koristi samo slojeve pažnje.

Na početku se audio zapis transformira u logaritamski mel-spektrogram izračunat Hammingovim prozorom od 25 ms na svakih 10 ms. Zatim se spektrogram dijeli u niz od N 16×16 manjih dijelova s preklapanjem od 6 u vremenskoj i frekvencijskoj dimenziji. Svaki od tih dijelova poravnava se u jednodimenzionalni oblik veličine 768 pomoću sloja linearne projekcije. Takvi dijelovi spektrograma predstavljaju ulaz u transformer. U transformeru se

koristi koder iz originalne transformerske arhitekture (Vaswani et al., 2017), bez dekodera. Na kraju, linearni sloj sa sigmoidnom aktivacijom preslikava reprezentaciju audio spektrograma u oznake za klasifikaciju (Gong et al., 2021).

Unaprijed naučeni model fino je podešen za detekciju srčanih šumova kroz 10 epoha, optimizatorom Adam, stopom učenja $3e-5$ i veličinom grupe 8. Postigao je odlične rezultate na skupu za testiranje: točnost od 92,52 %, preciznost od 80,68 %, odziv od 73,20 % i F1 mjeru od 76,76 %. Na Slika 5.1 prikazana je matrica zabune ovog modela. Iako nema bolji odziv od HuBERT-a, ovaj model postiže najbolji rezultat točnosti, preciznosti i F1 mjere od svih prijašnjih modela. Od 575 zapisa PCG-a u skupu za testiranje, pogrešno je klasificirao samo 43 zapisa.

5.4. Whisper

Whisper je jedan od najnovijih transformerskih modela koji pokazuje neke od najboljih rezultata u obradi zvuka. Najčešće je korišten u transkripciji i prevođenju govora, gdje dolazi dosta blizu ljudske razine sposobnosti i točnosti u prepoznavanju engleskog jezika. Razvili su ga istraživači tvrtke OpenAI 2022. godine (Radford et al., 2022).

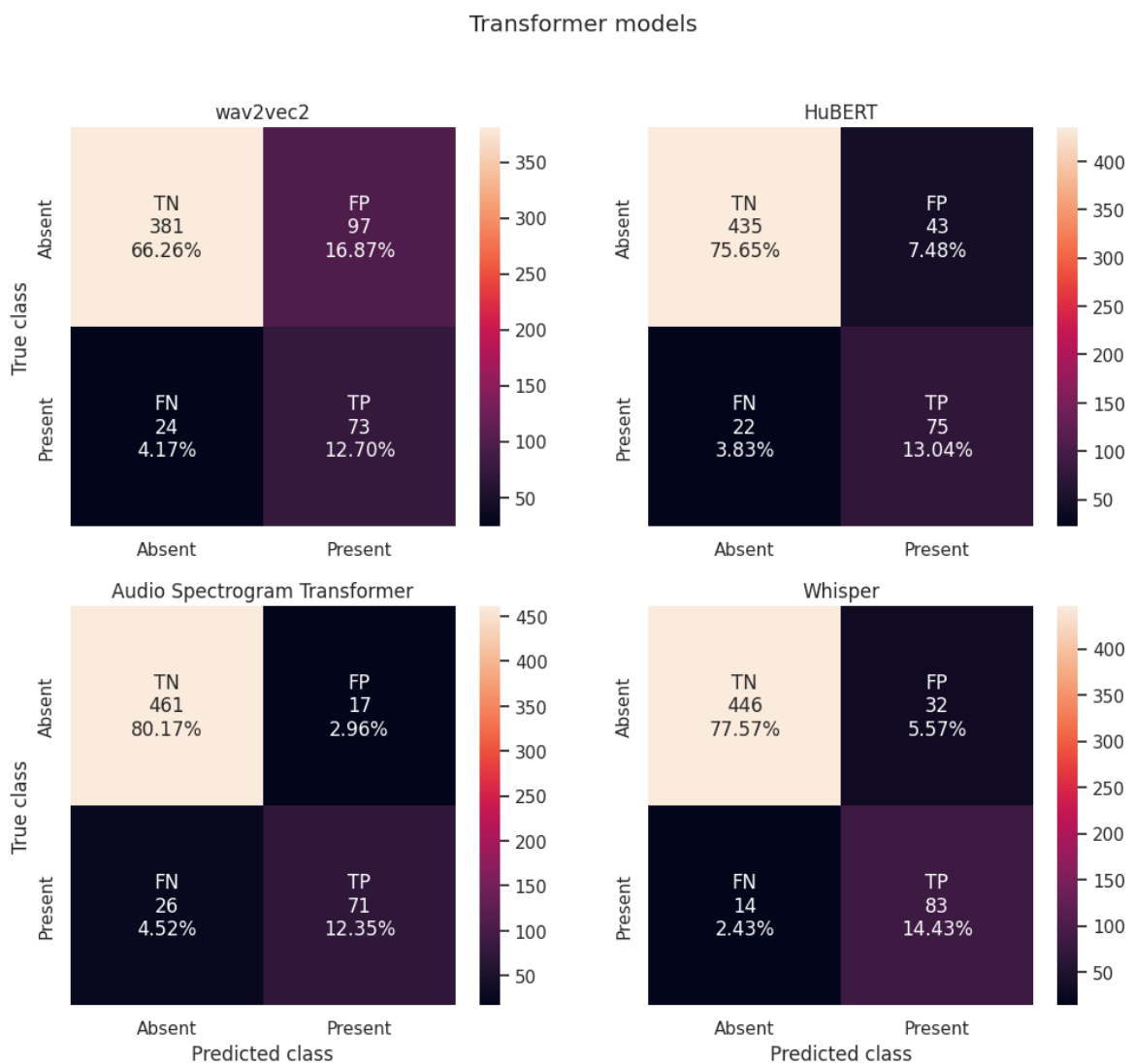
Svi zvučni zapisi naduzorkuju se na 16 kHz, a 80-kanalni prikaz mel-spektrograma logaritamske skale izračunava se na prozorima od 25 milisekundi s korakom od 10 milisekundi. Dva konvolucijska sloja obrađuju spektrograme i grade reprezentaciju audio zapisa koja se šalje na ulaz kodera. U arhitekturi za transkripciju govora korišteni su koder i dekodek slični originalnoj arhitekturi predstavljenoj u (Vaswani et al., 2017) (Radford et al., 2022).

S obzirom da je funkcija dekodera generiranje podataka, za zadatke klasifikacije taj dio modela nije potreban. Kod detekcije srčanih šumova, od arhitekture Whispera korišten je samo dekodek s dodatnim izlaznim slojem za klasifikaciju iz skrivenih stanja.

Unaprijed naučeni model učen je na skupu podataka LibriSpeech (Panayotov et al., 2015), te sadrži čak 242 milijuna parametara, što pokazuje ogroman kapacitet i kompleksnost ovog modela. Usporedbe radi, dvodimenzionalni konvolucijski model iz poglavlja 4.2.2 ima samo oko 3,3 milijuna parametara.

Ovaj model fino je podešen za detekciju srčanih šumova istim hiperparametrima kao i AST: optimizatorom Adam, 10 epoha učenja, stopom učenja $3e-5$ i veličinom grupe 8. Model

postiže točnost od 92,00 %, preciznost od 72,17 %, odziv od 85,57 % i F1 mjeru od 78,30 % na skupu za testiranje. Iako mu preciznost nije bolja od modela AST, odziv mu je bolji od HuBERT-a za 8,25%. Whisper na skupu za testiranje po rezultatima prestiže AST i stoji kao najuspješniji model dubokog učenja ovog istraživanja. Iako ima manju točnost i preciznost od AST-a, ipak postiže veću F1 mjeru i odziv koje su bitnije metrike za problem detekcije srčanih šumova. Na Slika 5.1 prikazana je matrica zabune ovog modela na skupu za testiranje.



Slika 5.1 Matrice zabune transformerskih modela

Tablica 3 Rezultati transformerskih modela

| <i>Transformer</i> | Točnost | Preciznost | Odziv | F1 mjera |
|--------------------|----------------|-------------------|--------------|-----------------|
| <i>Wav2vec 2.0</i> | 78,96 % | 42,94 % | 75,26 % | 54,68 % |
| <i>HuBERT</i> | 88,70 % | 63,56 % | 77,32 % | 69,77 % |
| <i>AST</i> | 92,52 % | 80,68 % | 73,20 % | 76,76 % |
| <i>Whisper</i> | 92,00 % | 72,17 % | 85,57 % | 78,30 % |

U Tablica 3 navedeni su rezultati svih transformerskih modela, koji su se općenito pokazali jako dobrim na zadatku detekcije srčanih šumova. Postižu bolje rezultate od bilo koje druge skupine modela ispitane u ovom istraživanju. To potvrđuje njihovu popularnost i laku sposobnost podešavanja raznim problemima.

Važno je napomenuti kako transformerski modeli imaju velik broj parametara i često je za njihovo učenje potrebno mnogo vremena i računalnih resursa pa je potrebno razmisliti isplati li ih se koristiti u svim primjenama.

6. Rasprava i rezultati

Tablica 4 Svi modeli istraživanja

| <i>Model</i> | Točnost | Preciznost | Odziv | F1 mjera |
|--------------------------------------|----------------|-------------------|----------------|-----------------|
| <i>Whisper</i> | 92,00 % | 72,17 % | 85,57 % | 78,30 % |
| <i>AST</i> | 92,52 % | 80,68 % | 73,20 % | 76,76 % |
| <i>Konvolucijsko-povratni model*</i> | 90,73 % | 73,40 % | 71,13 % | 72,25 % |
| <i>HuBERT</i> | 88,70 % | 63,56 % | 77,32 % | 69,77 % |
| <i>1D konvolucijski model*</i> | 85,84 % | 57,69 % | 61,86 % | 59,70 % |
| <i>Wav2vec 2.0</i> | 78,96 % | 42,94 % | 75,26 % | 54,68 % |
| <i>AlexNet*</i> | 81,49 % | 45,45 % | 46,39 % | 45,92 % |
| <i>2D konvolucijski model*</i> | 85,84 % | 68,18 % | 30,93 % | 42,55 % |
| <i>Naivni Bayesov klasifikator</i> | 76,70 % | 35,16 % | 42,38 % | 38,44 % |
| <i>ResNet50*</i> | 63,46 % | 26,07 % | 62,89 % | 36,86 % |
| <i>2D konvolucijski model</i> | 84,52 % | 60,53 % | 23,71 % | 34,07 % |
| <i>XGBoost</i> | 84,43 % | 65,91 % | 19,21 % | 29,74 % |
| <i>1D konvolucijski model</i> | 78,96 % | 32,86 % | 23,71 % | 27,54 % |
| <i>Konvolucijsko-povratni model</i> | 78,78 % | 31,34 % | 21,65 % | 25,61 % |
| <i>Logistička regresija</i> | 84,10 % | 66,67 % | 14,57 % | 23,91 % |
| <i>AlexNet</i> | 81,91 % | 40,54 % | 15,46 % | 22,39 % |

* označeni modeli učeni su metodom podjele na srčane cikluse

| | | | | |
|-----------------------------------|---------|---------|---------|---------|
| <i>Jednostavni povratni model</i> | 78,26 % | 26,67 % | 16,49 % | 20,38 % |
| <i>ResNet50</i> | 81,04 % | 34,21 % | 13,40 % | 19,26 % |
| <i>GRU model</i> | 79,30 % | 28,00 % | 14,43 % | 19,05 % |
| <i>Slučajna šuma</i> | 84,10 % | 86,67 % | 8,61 % | 15,66 % |
| <i>LSTM model</i> | 74,43 % | 15,28 % | 11,34 % | 13,02 % |

Tablica 4 prikazuje sve modele ispitane kroz ovo istraživanje sortirane po uspješnosti detekcije srčanih šumova. Modeli su sortirani po F1 mjeri koja je najvažnija metrika kod ovog skupa podataka zbog njegove nebalansiranosti. Sljedeća važna metrika nakon F1 mjere je odziv, koji je važan zbog medicinske prirode ovog problema. Odziv kažnjava lažno negativne primjerke kod klasifikacije. Kod sustava koji pomaže liječnicima u donošenju dijagnoza, bitno je smanjiti broj lažno negativnih klasifikacija, jer tada određeno prisutno oboljenje ostaje neopaženo.

Algoritmi strojnog učenja nisu se pokazali učinkovitima. Detekcija srčanih šumova iz snimki PCG-a očito je previše složen zadatak za ovakve algoritme, a kepralni koeficijenti mel-frekvencije na njihovom ulazu nisu dovoljni da bi se pronašao matematički prostor koji dobro razdvaja dvije klase PCG-a.

Konvolucijski modeli, koji inače dobro rade sa strukturiranim podacima, dali su zadovoljavajuće rezultate. Uspjeli su izlučiti neke korisne značajke iz audio zapisa i spektrograma, kako bi prepoznali srčani šum.

Uvođenjem metode podjele PCG zapisa na srčane cikluse i klasifikacije glasanjem za svaki srčani ciklus, uklonjen je dio podatkovnog šuma, smanjena pristranost većinskoj klasi u skupu podataka i najvažnije, znatno je povećana uspješnost modela.

Povratni modeli sami po sebi nisu uspjeli detektirati srčane šumove iz audio zapisa, ali njihovo korištenje zajednički s konvolucijskim slojevima poboljšalo je rezultate koje su već uspješni konvolucijski slojevi postizali. Tako je konvolucijsko-povratni model neuronske mreže treći po uspješnosti u ovom istraživanju, s točnošću od 90,73 % i F1 mjerom od 72,25 %.

Transformerski modeli pokazali su zašto su ovi nedavno izumljeni modeli sve popularniji u mnogim primjenama. Kroz par epoha finog podešavanja, lako su se prilagodili zadatku detekcije srčanih šumova, iako su bili unaprijed učeni na ljudskom govoru. Uspostavili su se kao najbolji modeli ovog istraživanja, uz manu da zahtijevaju znatno velike računalne i vremenske resurse, što je kompliciralo njihovo učenje. Najbolji model istraživanja je Whisper, koji je imao točnost od 92 %, **odziv od 85,57 %** i **F1 mjeru od 78,30 %**. AST je model koji je na drugom mjestu, sa samo malo lošijim rezultatima od Whispera, ali unatoč tome, ovaj model ima najveću vrijednost **točnosti od 92,52 %** i **preciznosti od 80,68 %**.

Budući da je skup podataka korišten u ovom istraživanju bio dio natjecanja The George B. Moody PhysioNet Challenge 2022 (Reyna et al., 2022), u literaturi postoji par znanstvenih radova koji koriste ovaj skup podataka i daju svoje rezultate na ovom ili sličnom problemu vezanom uz zapise PCG-a.

Istraživači A. Patwa, M. M. U. Rahman, T. Y. Al-Naffouri naučili su tri neuronske mreže u svrhu detekcije srčanih šumova. Jednodimenzionalnu konvolucijsku mrežu, LSTM mrežu i konvolucijsko-povratnu mrežu. Najvažniji dio njihovog rada bio je vremensko-frekvencijska analiza PCG segmenata korištenjem transformacije valičnog raspršenja. Njihov najbolji model bio je jednodimenzionalni konvolucijski, koji je postigao točnost od 87,05 % i F1 mjeru od 87,72 % (Patwa et al., 2023).

D. Vimalajeewa, C. Lee i B. Vidakovic predstavili su metodu otkrivanja srčanih šumova na temelju samo-sličnih svojstava. Samo-sličnost je karakterizirana procjenom fraktalnog ponašanja signala. Ova metoda evaluirana je s nekoliko standardnih klasifikatora strojnog učenja, a najbolji je stroj potpornih vektora koji postiže točnost od 76,61 % na skupu s balansiranim odnosom klasa (Vimalajeewa et al., 2022).

Ben Walker i suradnici naučili su dualnu Bayesovu ResNet mrežu gdje je svaki zapis pacijenta segmentiran u preklapajući log-mel-spektrogram. Također su koristili algoritam XGBoost kako bi iskoristili demografske podatke dostupne unutar skupa podataka. Njihov model natjecao se u detekciji srčanih šumova na natjecanju (Reyna et al., 2022), i postigao je utežanu točnost od 77,10% na skrivenom skupu za testiranje (Walker et al., 2022).

B. J. Singstad i suradnici su također razvili mrežu koja je sudjelovala u natjecanju (Reyna et al., 2022). Mreža je bila jednodimenzionalne arhitekture InceptionTime predstavljene u (Ismail Fawaz et al., 2020). Za detekciju srčanih šumova, postigla je utežanu točnost od 58,50 % (Singstad et al., 2022).

U radu L. Gua i njegovih suradnika ansambl jednodimenzionalne i dvodimenzionalne konvolucijske mreže postigao je točnost od 96,90 %. Takav ansambl također je korišten u radu za višestruko označavanje srčanih zvukova (Guo et al., 2022).

U usporedbi s drugim istraživačkim radovima koji su se bavili istim ili sličnim zadatkom, može se zaključiti kako su rezultati ovog rada među boljima. Naime, ovo istraživanje postiglo je bolje rezultate od većine radova iz ovog područja. S druge strane, postoje neki radovi koji su postigli bolje točnosti i F1 mjere, ali ovo istraživanje im je blizu po rezultatima.

Zaključak

Cilj ovog diplomskog rada bio je razvoj što korisnijeg i uspješnijeg sustava umjetne inteligencije za detekciju srčanih šumova iz snimki PCG-a. Iako je duboko učenje danas jako razvijeno, ovaj problem nije jednostavan za rješavanje. Za razvoj uspješnog modela, bilo je potrebno riješiti probleme nebalansiranosti i šuma u skupu podataka, izlučivanja korisnih značajki iz zvučnih zapisa, te zadovoljiti velike zahtjeve za računalnim resursima i vremenom.

Kroz ovo istraživanje ispitan je ukupno 21 model od kojih su 4 algoritma strojnog učenja, a 17 su modeli dubokog učenja.

Od algoritama strojnog učenja ispitan su logistička regresija, naivni bayesov klasifikator, slučajna šuma i XGBoost, koji su dali lošije rezultate od dubokih modela. Ovi algoritmi radili su s reprezentacijama zvuka preko kepstralnih koeficijenata mel frekvencije, koji se baziraju na Fourierovoj transformaciji.

Od konvolucijskih modela ispitan su jednodimenzionalni model koji je obrađivao čisti zvuk, dvodimenzionalni konvolucijski model koji je obrađivao mel-spektrograme i unaprijed naučeni modeli AlexNet i ResNet50. Konvolucijski modeli dali su osrednje rezultate.

Naučeni povratni modeli bili su jednostavni povratni model, LSTM model i GRU model. Oni sami po sebi nisu davali dobre rezultate, ali su u kombinaciji s konvolucijskim mrežama poboljšali rezultate.

Predstavljena je metoda učenja modela podjelom zapisa PCG-a na srčane šumove i uvođenjem glasanja kod donošenja odluke o prisutnosti šuma. Ta je metoda znatno poboljšala rezultate konvolucijskih modela, od kojih je najbolji bio konvolucijsko-povratni model. Taj model treći je po uspješnosti u ovom diplomskom radu.

Ispitani su također i najnoviji modeli u području dubokog učenja koji su transformerske arhitekture. Uzeti su unaprijed naučeni modeli koji su fino podešeni za detekciju srčanih šumova. Modeli su wav2vec 2.0, HuBERT, AST i Whisper. AST postiže najveću točnost od **92,52%** i najveću preciznost od **80,68%**. Unatoč tome, Whisper je najbolji model ovog diplomskog rada s točnošću od 92,00%, preciznosti od 72,17%, odzivom od **85,57%** i F1 mjerom od **78,30%**.

U usporedbi s drugim radovima na istom skupu podataka i na istom ili sličnom zadatku, ima i boljih i lošijih rezultata, ali modeli ovog rada definitivno stoje među boljima.

Mogućnost poboljšanja učinkovitosti modela umjetne inteligencije uvijek postoji. Što se tiče ovog istraživanja, moguće je unaprijediti postupke predobrade zvučnih zapisa s više ekspertnog znanja iz obrade zvuka i medicine. Optimizacijom hiperparametara modela, koju nije bilo vremena provoditi, također bi se mogli poboljšati rezultati. Unatoč tome, ovaj rad donosi usporedbu velikog broja fundamentalnih modela strojnog i dubokog učenja i predstavlja modele s visokom razinom učinkovitosti detekcije srčanih šumova iz snimki fonokardiograma.

Literatura

- Mattingly, Q., 2024., *Cardiovascular diseases*, World Health Organization, <https://www.who.int/health-topics/cardiovascular-diseases>, pristupljeno 10. siječnja 2024.
- World Health Organization (WHO), World health statistics 2023: monitoring health for the SDGs, Sustainable Development Goals. Geneva: World Health Organization; 2023.
- Oliveira, J., Renna, F., Costa, P.D., Nogueira, M., Oliveira, C., Ferreira, C., Jorge, A., Mattos, S., Hatem, T., Tavares, T., Elola, A. The CirCor DigiScope dataset: from murmur detection to murmur classification. *IEEE journal of biomedical and health informatics*. 26, 6, (2021), str. 2524-2535.
- Subasi A. Practical guide for biomedical signals analysis using machine learning techniques: A MATLAB based approach. Academic Press, 2019.
- Libby, P., Bonow, R., Mann, D., Zipes, D. Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine. 8th edition. Elsevier Science, 2007.
- Gheorghe, A., Griffiths, U., Murphy, A., Legido-Quigley, H., Lamptey, P., Perel, P. The economic burden of cardiovascular disease and hypertension in low-and middle-income countries: a systematic review. *BMC public health*. 18, 1, (2018), str. 1-11
- Murphy, A., Palafox, B., Walli-Attaei, M., Powell-Jackson, T., Rangarajan, S., Alhabib, K.F., Calik, K.B., Chifamba, J., Choudhury, T., Dagenais, G., Dans, A.L. The household economic burden of non-communicable diseases in 18 countries. *BMJ global health*. 5, 2, (2020)
- Leal, J., Luengo-Fernández, R., Gray, A., Petersen, S., Rayner, M. Economic burden of cardiovascular diseases in the enlarged European Union. *European heart journal*. 27, 13, (2006), str. 1610-1619.
- Soler-Soler, J., Galve, E. Worldwide perspective of valve disease. *Heart*. 83, 6, (2000), str. 721-725.
- Burns, J., Ganigara, M., Dhar, A. Application of intelligent phonocardiography in the detection of congenital heart disease in pediatric patients: a narrative review. *Progress in Pediatric Cardiology*. 64, (2022), str. 101455.
- Ranganathan, N., Sivaciyan, V., Saksena, F.B. The art and science of cardiac physical examination: with heart sounds and pulse wave forms on CD. Springer Science & Business Media, 2007.
- Baguet, J.P., Barone-Rochette, G., Tamisier, R., Levy, P., Pepin, J.L. Mechanisms of cardiac dysfunction in obstructive sleep apnea. *Nature Reviews Cardiology*. 9, 12, (2012), str. 679-688.
- Gray, H. *Anatomy of the Human Body*. Philadelphia: Lea & Febiger, 1918; New York: Bartleby.com, 2000. <https://www.bartleby.com/lit-hub/anatomy-of-the-human-body>. [Online]
- Oweis, R.J., Hamad, H., Shammout, M. Heart sounds segmentation utilizing teager energy operator. *Journal of Medical Imaging and Health Informatics*. 4, 4, (2014), str. 488-499.

- Dornbush, S., Turnquest, A.E. Physiology, Heart Sounds. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2019. Dostupno na: <https://www.ncbi.nlm.nih.gov/books/NBK541010/>
- Rangayyan, R. M. Biomedical signal analysis. Izdanje 33. New Jersey: John Wiley & Sons, 2015.
- Pollock, J.D., Makaryus, A.N. Physiology, Cardiac Cycle. In: StatPearls. Treasure Island (FL): StatPearls Publishing, 2020.
- Gomes, E., Bentley, P., Coimbra, M., Pereira, E., Deng, Y. Classifying heart sounds: Approaches to the PASCAL challenge. HEALTHINF 2013 - Proceedings of the International Conference on Health Informatics, (2013), str. 337–340.
- Liu, C., Springer, D., Li, Q., Moody, B., Juan, R.A., Chorro, F.J., Castells, F., Roig, J.M., Silva, I., Johnson, A.E., Syed, Z. An open access database for the evaluation of heart sound algorithms. *Physiological measurement*. 37, 12, (2016), str. 2181.
- Spadaccini, A., Beritelli, F. Performance evaluation of heart sounds biometric systems on an open dataset. In 2013 18th International Conference on Digital Signal Processing (DSP), (2013), str. 1-5.
- Oliveira, J., Renna, F., Mantadelis, T., Coimbra, M. Adaptive sojourn time HSMM for heart sound segmentation. *IEEE journal of biomedical and health informatics*. 23, 2, (2018), str. 642-649.
- Semmlow, J.L. Biosignal and biomedical image processing: MATLAB-based applications. Marcel Dekker, Inc., 2004.
- Giannakopoulos, T., Pikrakis, A. Introduction to audio analysis: a MATLAB® approach. Academic Press, 2014.
- Raschka, S. Python Machine Learning. Packt Publishing Ltd, 2015.
- Belyadi, H., Haghghat, A. Machine Learning Guide for Oil and Gas Using Python - A Step-by-Step Breakdown with Data, Algorithms, Codes, and Applications. Gulf Professional Publishing, 2021.
- Subasi A. Practical machine learning for data analysis using python. Academic Press, 2020.
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H. Top 10 algorithms in data mining. *Knowledge and information systems*. 14, (2008), str. 1-37.
- Quinlan, J.R. Induction of decision trees. *Machine learning*. 1, (1986) str. 81-106.
- Chen, T., Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, (2016), str. 785-794.
- Jović A., Dubinska analiza podataka, materijali s predavanja, FER, 2023.
- Bengio, Y., Goodfellow I., Courville, A. Deep learning. Prvo izdanje. Cambridge, MA, USA: MIT press, 2017.
- Buduma, N., Buduma, N., Papa, J. Fundamentals of deep learning. O'Reilly Media, Inc., 2022.

- Krizhevsky, A., Sutskever, I., Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 25, (2012).
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, (2009), str. 248-255.
- He, K., Zhang, X., Ren, S., Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), str. 770-778.
- Hochreiter, S., Schmidhuber, J. Long short-term memory. *Neural computation*. 9, 8, (1997), str. 1735-1780.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*. (2014).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*. 30, (2017).
- Amatriain X. Transformer models: an introduction and catalog. *arXiv preprint arXiv:2302.07730*. (2023).
- Baevski, A., Zhou, Y., Mohamed, A., Auli, M. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*. 33, (2020), str. 12449-12460.
- Panayotov, V., Chen, G., Povey, D., Khudanpur, S. Librispeech: an ASR corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, (2015), str. 5206-5210.
- Hsu, W.N., Bolte, B., Tsai, Y.H., Lakhota, K., Salakhutdinov, R., Mohamed, A. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 29, (2021), str. 3451-3460.
- Gong, Y., Chung, Y.A., Glass, J. Ast: Audio spectrogram transformer. *arXiv preprint arXiv:2104.01778*. (2021).
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. (2020).
- Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, (2023), str. 28492-28518.
- Reyna, M.A., Kiarashi, Y., Elola, A., Oliveira, J., Renna, F., Gu, A., Alday, E.A., Sadr, N., Sharma, A., Mattos, S., Coimbra, M.T. Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. In *2022 Computing in Cardiology (CinC)*, (2022), izdanje 498, str. 1-4.
- Patwa, A., Rahman, M.M., Al-Naffouri, T.Y. Heart Murmur and Abnormal PCG Detection via Wavelet Scattering Transform & a 1D-CNN. *arXiv preprint arXiv:2303.11423*. (2023).

Vimalajeewa, D., Lee, C., Vidakovic, B. A Method for Detecting Murmurous Heart Sounds based on Self-similar Properties. arXiv preprint arXiv:2306.05283. (2023).

Walker, B., Krones, F., Kiskin, I., Parsons, G., Lyons, T., Mahdi, A. Dual Bayesian ResNet: A Deep Learning Approach to Heart Murmur Detection. In 2022 Computing in Cardiology (CinC), (2022), izdanje 498, str. 1-4.

Singstad, B.J., Gitau, A.M., Johnsen, M.K., Ravn, J., Bongo, L.A., Schirmer, H. Phonocardiogram classification using 1-dimensional inception time convolutional neural networks. In 2022 Computing in Cardiology (CinC), (2022), izdanje 498, str. 1-4.

Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, DF., Weber, J., Webb, G.I., Idoumghar, L., Muller, PA., Petitjean, F. InceptionTime: Finding AlexNet for time series classification. Data Mining and Knowledge Discovery. 34, 6, (2020), str:1936-1962.

Guo, L., Davenport, S., Peng, Y. Deep CardioSound-An Ensembled Deep Learning Model for Heart Sound MultiLabelling. arXiv preprint arXiv:2204.07420. (2022).

Sažetak

Naslov:

Detekcija srčanih šumova iz snimki fonokardiograma metodama dubokog učenja

Sažetak:

Urođene srčane bolesti stvaraju velik teret za zdravstveni sustav i odnose mnoge ljudske živote. Srčani šumovi često mogu ukazivati na neku takvu bolest. Fonokardiogram je zapis zvuka srca snimljen digitalnim stetoskopom. Računalnom analizom fonokardiograma i detekcijom srčanih šumova može se pomoći liječnicima kod dijagnoza srčanih bolesti. U ovom radu su ispitana i uspoređena 4 modela strojnog i 17 modela dubokog učenja na temelju uspješnosti detekcije srčanih šumova na skupu podataka CirCor DigiScope. Naučeni su modeli koji detektiraju srčane šumove s točnosti od 92 %, odzivom od 85 % i F1 mjerom od 78 %.

Ključne riječi:

Fonokardiogram, srčani šum, srčani ciklus, strojno učenje, duboko učenje, Fourierova transformacija, spektrogram, konvolucijska mreža, povratna mreža, transformer

Summary

Title:

Detection of heart murmurs from phonocardiogram recordings using deep learning methods

Summary:

Congenital heart diseases create a great burden on the healthcare system and take many human lives. Heart murmurs can often indicate such a disease. A phonocardiogram is a recording of heart sounds taken with a digital stethoscope. Computer analysis of phonocardiograms and detection of heart murmurs can help doctors diagnose heart diseases. In this paper, 4 machine learning models and 17 deep learning models were tested and compared based on the success of heart murmur detection on the CirCor DigiScope dataset. Models were trained to detect heart murmurs with an accuracy of 92 %, recall of 85 % and F1 score of 78 %.

Keywords:

Phonocardiogram, heart murmur, cardiac cycle, machine learning, deep learning, Fourier transform, spectrogram, convolutional network, recurrent network, transformer

Skraćenice

| | | |
|-------|-------------------------------------|---|
| PCG | <i>Phonocardiogram</i> | Fonokardiogram |
| FT | <i>Fourier transform</i> | Fourierova transformacija |
| CAD | <i>Coronary Artery Disease</i> | Bolest koronarnih arterija |
| VHD | <i>Valvular heart disease</i> | Bolest srčanih zalistaka |
| CHD | <i>Congenital heart disease</i> | Prirođene srčane mane |
| MFCCs | Mel-frequency cepstral coefficients | Kepstralni koeficijenti mel-frekvencije |
| MSE | <i>Mean squared error</i> | Srednja kvadratna pogreška |
| MLE | <i>Maximum likelihood estimator</i> | Procjena maksimalne izglednosti |
| LSTM | <i>Long short-term memory</i> | Duga kratkoročna memorija |
| GRU | <i>Gated recurrent unit</i> | Propusna povratna ćelija |