

Širdies plakimo garso įrašų (fonokardiogramų) klasifikavimui skirtų metodų tyrimas

Mantas Gricius

Kauno technologijos universitetas,
K. Donelaičio g. 73, LT-44249 Kaunas
mantas.gricius@ktu.edu

Santrauka. Pasaulio sveikatos organizacijos duomenimis, širdies ir kraujagyslių ligos yra pagrindinė mirties priežastis pasaulyje. Dirbtinio intelekto klasifikavimo algoritmai gali padėti aptikti širdies ligas ankstyvoje stadijoje. Šio darbo tikslas – sukurti algoritmą fonokardiogramų klasifikavimui (tipinis / netipinis signalas) ir ištirti duomenų įtaką klasifikavimo rezultatams. Viena pagrindinių problemų šios srities sprendimuose yra ribotas apmokymo duomenų kiekis. Dažniausiai šiuose tyrimuose naudojamas duomenų rinkinys yra „2016 CinC/PhysioNet iššūkiu“ duomenų bazė. Tačiau šiame duomenų rinkinyje egzistuoja duomenų disbalansas – tipinių garso signalų yra beveik 4 kartus daugiau nei netipinių. Todėl tyrime išbandomi 3 skirtingi klasifikavimo modeliai (atraminių vektorių klasifikatorius (SVM), atsitiktiniai miškai, konvoliucinis neuroninis tinklas (CNN)) ir palyginama, kaip duomenų disbalansas paveikia klasifikavimo tikslumą. SVM klasifikavo netipinius signalus prasčiausiai (jautrumas – 0,6683), o geriausiai klasifikavo CNN modelis (MAcc – 0,8470). Duomenų disbalansui pašalinti buvo atliekama netipinių signalų augmentacija ir sukurti sintetiniai duomenys, kurie išlygino įrašų kiekį kiekvienai klasei. Pašalinus duomenų disbalansą, visi modeliai klasifikavo netipinius garso signalus ženkliai tiksliau, SVM pasiekė net 20% geresnį jautrumo įvertį. O tipinių signalų klasifikavimo tikslumas suprastėjo maždaug 5%.

Raktiniai žodžiai: CNN, GAN, širdies garso įrašai, klasifikavimas, dirbtinių neuronų tinklas, mašininis mokymas.

1 Įvadas

Pasaulio sveikatos organizacijos duomenimis, širdies ir kraujagyslių ligos yra pagrindinė mirties priežastis pasaulyje [1]. Širdies vožtuvų ligos (angl. valvular heart diseases, VHD) yra viena iš pagrindinių širdies ir kraujagyslių ligų priežasčių, dėl kurių visame pasaulyje yra didelis mirtingumas [2].

Ankstyvas tokių problemų diagnozavimas gali padėti sumažinti mirtingumą sustabdant ligos vystymąsi.

Šio darbo tikslas – sukurti algoritmą fonokardiogramų (angl. phonocardiogram, PCG) klasifikavimui (tipinis / netipinis signalas) ir ištirti duomenų įtaką klasifikavimo rezultatams. Sukurto algoritmo ir egzistuojančių sprendimų rezultatų vientisumui užtikrinti, naudojamas „2016 CinC/PhysioNet iššūkių“ duomenų rinkinys [3]. Tikslui pasiekti atliekami eksperimentai naudojant skirtingus mašininio mokymosi (angl. machine learning, ML) modelius. „2016 CinC/PhysioNet iššūkių“ duomenų rinkinyje širdies plakimo garso įrašai yra suskirstyti į tipinius ir netipinius, tačiau tipinių fonokardiogramų yra gerokai daugiau (2575) nei netipinių (665). Toks duomenų disbalansas sukuria klasifikavimo modelio šališkumą tipiniams garso įrašams. Dėl šios priežasties naudojami papildomi garsų augmentacijos metodai, tokie kaip Gauso triukšmas (angl. Gaussian noise), laiko tempimas (angl. time stretch), tono keitimas (angl. pitch shift), garso keitimas (angl. shift) bei generatyvusis priešininkų tinklas (angl. generative adversarial network, GAN) sintetiniams duomenims kurti. Tai turėtų padėti pašalinti šališkumą tipiniams signalams, efektyviau apmokyti modelius ir gauti tikslesnius klasifikavimo rezultatus.

2 Egzistuojančių fonokardiogramų klasifikavimo sprendimų ir literatūros analizė

Literatūroje matoma tendencija, jog pasirinktas duomenų rinkinys „Physionet Challenge 2016“ iš ties gausiai paplitęs šios srities moksliniuose darbuose [2, 4, 5, 6, 7] ir kadangi duomenų rinkinys turi labai didelę įtaką klasifikavimo tikslumui, naudodami tą patį duomenų rinkinį galėsime geriau palyginti gautus rezultatus. Požymių išskyrimui naudojami labai įvairūs ir skirtingi metodai. Krishnan et al., [4], 2020 ir Alkhodari et al., [2], 2021 naudojo 3-jų sluoksnių vienmatį konvoliucinį neuroninį tinklą (angl. convolutional neural network, CNN); Shuvo et al., [6], 2021 taip pat naudojo CNN, tačiau 4-ių sluoksnių; tuo tarpu Zabihi et al., [5], 2016 išskaičiavo požymius pasitelkiant 5 skirtingus garso signalų apdorojimo metodus; o Demir et al., [8], 2019 pasitelkė net iš anksto apmokytus giliuosius neuroninius tinklus (angl. pretrained deep neural networks, DNN) VGG 16, VGG 19 ir AlexNet. Analizėje matome, jog dažniausiai naudojamas požymių išskyrimo metodas buvo kelių sluoksnių konvoliucinis neuroninis tinklas. Tyrimų rezultatams tarpusavyje palyginti naudojamas balansuotas tikslumas (angl. bal-

anced accuracy, MAcc). Tyrimai taip pat naudojo skirtingus būdus ir pačiam fonokardiogramų klasifikavimui. Krishnan et al., [4], 2020 ir Zabihi et al., [5], 2016 pasitelkė tiesioginio sklaidimo neuroninius tinklus (angl. feed-forward neural networks, FNN) ir atitinkamai pasiekė 0,8574 ir 0,8590 MAcc. Alkhodari et al., [2], 2021 naudojo dvikryptį ilgalaikės trumpalaikės atminties (angl. bi-directional long short-term memory, BiLSTM), rekurentinį neuroninį tinklą (angl. recurrent neural network, RNN) ir pasiekė 0,8731 MAcc. O geriausią rezultatą pasiekti pavyko Chen et al., [7], 2020 – 0,9391 MAcc. Iš atliktos analizės galime matyti pagrindines širdies plakimo garso įrašų klasifikavimo darbų tendencijas: dauguma darbų modelio apmokymui, validacijai ir testavimui naudoja „Physionet Challenge 2016“ duomenų rinkinį; gan dažnai pasirenkama naudoti konvoliucinius neuroninius tinklus požymių išskyrimui; klasifikavimui dažnai naudojami įvairūs dirbtinių neuronų tinklai (FNN, RNN, CNN). Taip pat verta atsižvelgti ir į esamų sprendimų silpnąsias savybes ir didžiausius iššūkius. Chen et al., [9], 2021 atliko išsamią fonokardiogramų klasifikavimo darbų analizę ir aptiko pagrindines problemas ir iššūkius:

- Ribotas duomenų kiekis apmokymui;
- Apmokymo efektyvumas ir sparta;
- Giliesiems dirbtiniams neuroniniams tinklams apmokyti reikalingi atminties ir skaičiavimo ištekliai.

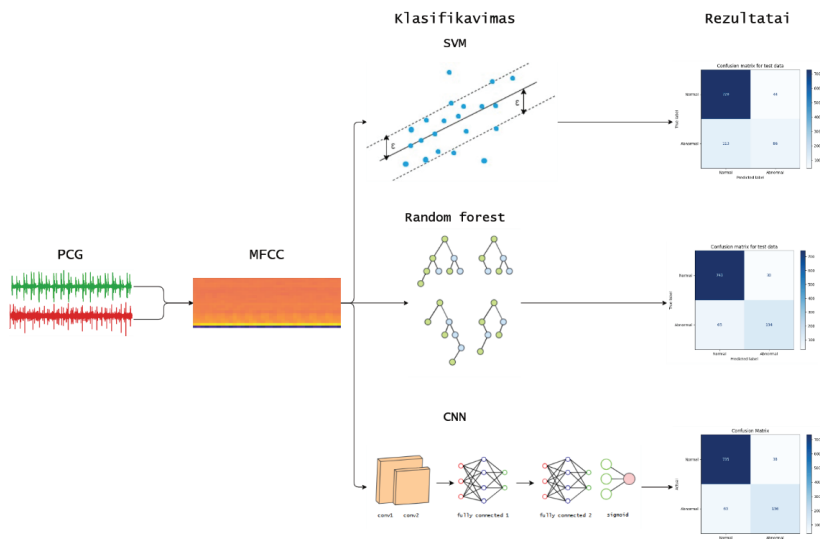
Atsižvelgiant į šiuos iššūkius galime orientuotis į sritis, kurias galima patobulinti ir galimai gauti geresnius rezultatus arba atrasti kaip palengvinti patį klasifikavimo procesą.

3 Eksperimentų eiga

Prieš pradėdant eksperimentus vienas labai svarbus žingsnis yra tinkamai atskirti ir apdoroti duomenis. Kadangi duomenų kiekis nėra labai didelis (3240 įrašai), norint atlikti tikslesnį duomenų testavimą pasirinkta skirti 30% duomenų (972 įrašai) testavimui ir 70% duomenų (2268 įrašai) modelio apmokymui. Eksperimentams svarbu ne tik duomenų kiekis, tačiau ir jų kokybė. Peržvelgus duomenis buvo pastebėtas duomenų disbalansas. Iš 2268 įrašų tik 457 įrašai apmokymo imtyje buvo netipiniai, tai reiškia, kad tipinių širdies plakimo garso įrašų apmokymo imtyje yra beveik 4 kartus daugiau nei netipinių. Tai daugumai modelių sukuria šališkumą ir didesnę

tikimybę, kad garsai bus atpažinti kaip tipiniai, kas gali nulemti labai prastą netipinių fonokardiogramų klasifikavimą. Toks duomenų disbalansas suteikia progą palyginti kaip skirtingi modeliai susitvarko su šia problema ir kaip šios problemos pašalinimas įtakoja tiek tipinių, tiek netipinių širdies plakimo garsų klasifikavimą. Požymių išskyrimas yra pirmas žingsnis šios srities tyrimų eksperimentuose. Šiame darbe požymių išskyrimui naudojami Mel cepstrum koeficientai (angl. Mel Frequency Cepstrum Coefficients, MFCC). Pasirinkta naudoti pirmuosius 20 MFCC koeficientų, nes tolimesni koeficientai gali lemti modelio persimokymą. Duomenų rinkinyje garsai yra skirtingos trukmės (nuo 5 iki 120 sekundžių) ir prieš požymių išskyrimą nėra niekaip apkarpmi ar apdorojami siekiant išlaikyti kuo daugiau informacijos. Dėl skirtingos trukmės įrašų, gaunamas MFCC kadrų skaičius kiekvienam koeficientui skiriasi priklausomai nuo garso trukmės. Daugumai modelių naudoti skirtingų matmenų įvesties duomenis nėra priimtina, todėl modelių apmokymui ir testavimui naudojami kiekvieno MFCC koeficiento kadrų vidurkiai, taigi kiekvienas duomenų įvesties įrašas turi lygiai 20 savybių. Eksperimentams pasirinkti naudoti 3 skirtingi modeliai (žr. 1 pav.): atraminių vektorių klasifikatorius (angl. support vector machine, SVM) pasirinktas kaip pirminis modelis, dėl populiarumo šios srities tyrimuose; atsitiktinių sprendimų miškų (angl. random forest) modelis pasirinktas, kadangi dėl savo prigimties ir veikimo būdo gali nesukurti šališkumo ir geriau atlikti klasifikavimą esant duomenų disbalansui; 2-jų sluoksnių CNN modelis dėl literatūroje pasiektų rezultatų. Pasirinkus kelis skirtingus modelius galėsime palyginti, kaip duomenų disbalansas įtakoja modelių klasifikavimo tikslumą. Kiekvienas modelis taip pat turi parametrus, nuo kurių priklauso klasifikavimo tikslumas. SVM modeliui pasirinkta naudoti radialinės bazinės funkcijos branduolį (angl. Radial Basis Function kernel, RBF), dėl gebėjimo geriau prisitaikyti ir klasifikuoti netiesinius duomenis. Atsitiktinių miškų algoritmui pasirinkta naudoti 50 sprendimų medžių siekiant užtikrinti pusiausvyrą tarp modelio tikslumo ir skaičiavimų efektyvumo. Minimalus įrašų skaičius medžio padalijimui pasirinktas numatytasis parametras – 2 įrašai, tai užtikrina, kad medžiai būtų pakankamai detalūs ir užfiksuotų sunkiai aptinkamas savybes. Minimalus įrašų skaičius lapuose (galutinėse medžių viršūnėse) pasirinktas taip pat 2 įrašai, siekiant šiek tiek sumažinti persimokymo (angl. overfitting) riziką, tačiau vis dar išlaikant gan aukštą medžio detalumą. Kadangi atliekama binarinė klasifikacija (šir-

dies garsai klasifikuojami į tipinius ar netipinius) CNN pasirinkta naudoti binarinės kryžminės entropijos (angl. Binary Cross Entropy, BCE) nuostolių funkciją. Kiti CNN parametrai reikalavo šiek tiek eksperimentavimo. Išbandžius keletą skirtingų optimizatorių, pasirinktas Adafactor optimizatorius ir 0.01 mokymosi greitis. Pasirinkus parametrus ir atlikus eksperimentus, tolimesniu žingsniu, siekiama įvertinti, kaip modelio rezultatai kinta naudojant subalansuotą duomenų rinkinį, kuriame abiejų klasių įrašų skaičius yra vienodas ar bent jau panašus. Tam reikia arba sumažinti tipinių garso įrašų kiekį, arba padidinti netipinių įrašų kiekį. Kadangi duomenų kiekis nėra labai didelis, jį sumažinus būtų sudėtinga sėkmingai apmokyti modelį, taigi reikia padidinti netipinių įrašų kiekį. Šiam tikslui pasiekti atliekama netipinių garso įrašų augmentacija pasitelkiant Gauso triukšmą, laiko tempimą, tono keitimą, garso keitimą. Gauso triukšmas suteikia papildomą triukšmą garso įrašams ir imituoja realistinę aplinką (pvz., kūno judesiai, elektroninio stetoskopo triukšmas). Tai padeda modeliui geriau prisitaikyti prie triukšmingų signalų ir tiksliau atpažinti netipinių širdies garsų savybes. Laiko tempimas padeda imituoti įvairius netipinius širdies garsus su skirtingais širdies ritmais (lėtesnis – bradikardija, greitesnis – tachikardija). Tono keitimas simuliuoja skirtingų pacientų anatomiją ir stetoskopo padėtį – dėl kūno masės ar padėties širdies garsai gali turėti aukštesnį arba žemesnį toną, tokių garsų įvairovė leidžia modeliui mažiau priklausyti nuo įvairių dažnių ir labiau atkreipti dėmesį į garsų laiką ir struktūrą (pvz. užesius, papildomus garsus). Garso keitimas perstumia garso signalą laike, todėl modelis neprisiriša prie garsų vietos įrašė. Tai padeda modeliui atpažinti reikšmingas savybes bet kurioje įrašo vietoje ir būti nepriklausomam nuo įrašo trukmės. Pasitelkiant šiuos 4 augmentacijos metodus ir juos atsitiktiniu būdu kombinuojant, kiekvienam netipiniam garso įrašui sukuriamas po dar 3 sintetinius garso įrašus bei taip subalansuojamas apmokymo duomenų rinkinys. Prijungus sintetinius duomenis prie apmokymo duomenų, gaunamas duomenų rinkinys turi 1811 tipinių širdies plakimo garso įrašų ir 1828 netipinių. Tuomet naudojant papildytą duomenų rinkinį vykdoma ta pati 1 pav. pavaizduota veiksmų seka. Atlikus klasifikavimą kiekvienu atveju apskaičiuojama klaidų matrica, kurios pagalba galima apskaičiuoti ir kitas modelio vertinimo metrikas (tikslumą, jautrumą, F įvertį, specifiškumą).



1 pav. Eksperimentų eigos schema.

4 Rezultatai

Atliktų eksperimentų rezultatai atvaizduoti 1 ir 2 lentelėse. Modelio tikslas tyrime yra suklasifikuoti gautą garso signalą kaip tipinį arba netipinį širdies plakimą. Iš loginės ir probleminės svarbos, svarbiau, kad būtų teisingai klasifikuojami netipiniai širdies plakimo garso signalai, nes šiuo atveju būtų galima įspėti pacientą, kad jam ištis gresia pavojus ir reikia kardiologo konsultacijos. Todėl apžvelgiant rezultatus daugiausiai skirsime dėmesio jautrumo įverčiui, kuris reprezentuoja kiek netipinių širdies plakimo garsų buvo suklasifikuota teisingai. 1 lentelėje galime matyti kokią įtaką turi modelio pasirinkimas apmokant modelį klasifikavimui, kai duomenų rinkinys yra nesubalansuotas ir dominuoja vienos klasės įrašai. Matome, jog SVM modelis labiausiai nukentėjo nuo duomenų disbalanso ir klasifikavo tipinius signalus itin gerai (specifiškumas – 0,9777), o netipinius itin blogai (jautrumas – 0,6683). Nors ir perėjus prie atsitiktinių miškų modelio rezultatai neturėjo ženklų pokyčių, galime pastebėti, jog modelio prigimtis leido šiam modeliui ne taip stipriai prisirišti prie dominuojančios klasės ir klasifikuoti netipinius signalus maždaug 1,5% tiksliau. Na o geriausius rezultatus pasiekti pavyko naudojant CNN modelį: MAcc – 0,8470, jautrumas - 0,7307. Taip pat galime

pastebėti, kad kai modeliai tiksliau klasifikuoja netipinius signalus, tipinių signalų klasifikavimo tikslumas šiek tiek suprastėja (žr. 1 lentelės specifishkumo stulpelį).

1 lentelė. Rezultatai naudojant originalius duomenis (be augmentacijos).

MODELIS	TIKSLUMAS	F1 ĮVERTIS	SPECIFIŠKUMAS	JAUTRUMAS	MAcc
SVM	0,9115	0,7637	0,9777	0,6683	0,8230
RANDOM FOREST	0,9104	0,7654	0,9725	0,6826	0,8276
CNN	0,9135	0,7835	0,9633	0,7307	0,8470

2 lentelėje pateikti modelių klasifikavimo rezultatais kai apmokymo imtyje naudojami sintetiniai duomenys išlyginti įrašų kiekį kiekvienai klasei ir panaikinamas duomenų disbalansas. Galima pastebėti, jog subalansavus duomenų rinkinį visi modeliai ženkliai tiksliau klasifikuoja netipinius garsus. SVM jautrumas pagerėjo net beveik 20% ir kadangi atsitiktinių miškų modelis jau nebeturi pranašumo dėl duomenų disbalanso, SVM pranoksta atsitiktinių miškų modelį klasifikuojant ir netipinius širdies plakimo garso signalus. Tačiau taip pat matomas skirtumas ir 2 lentelės specifishkumo stulpelyje. Nors pokytis ne toks ženklus, pašalinus duomenų disbalansą, modeliai maždaug 5% prasčiau klasifikuoja tipines fonokardiogramas.

2 lentelė. Rezultatai naudojant duomenų rinkinį papildytą sintetiniais duomenimis.

MODELIS	TIKSLUMAS	F1 ĮVERTIS	SPECIFIŠKUMAS	JAUTRUMAS	MAcc
SVM	0,9002	0,7810	0,9188	0,8317	0,8753
RANDOM FOREST	0,8909	0,7644	0,9083	0,8269	0,8676
CNN	0,9135	0,8090	0,9293	0,8557	0,8925

5 Išvados

Nors ir „2016 CinC/PhysioNet iššūkio“ duomenų rinkinys sudarytas iš moksliskai patvirtintų ir tikrai kokybiškų duomenų, deja, tipinių garso signalų kiekis yra gerokai didesnis nei netipinių, kas sukuria duomenų disbalansą. Taigi ribotas duomenų kiekis apmokymui vis dar lieka viena iš esminių problemų šio tipo tyrimuose. Šiame darbe buvo įvertinta duomenų disbalanso įtaka skirtingiems dirbtinio intelekto klasifikavimo metodams ir kaip pasikeičia tokių dirbtinio intelekto modelių klasifikavimo tikslumas pašalinus

duomenų disbalanso problemą. Duomenų disbalansas didžiausią įtaką turėjo SVM klasifikavimo modeliui, kuris pasiekė tik 0,6683 jautrumą, atsitiktinių medžių algoritmas klasifikavo netipinius širdies plakimo garso signalus maždaug 1,5% tiksliau (jautrumas – 0,6826), o CNN klasifikavo net 9% tiksliau (jautrumas – 0,7307). Duomenų disbalansui pašalinti buvo atlikta augmentacija netipiniams garsams ir sukurta sintetinių duomenų, įrašų kiekiui kiekvienai klasei išlyginti. Naudojant duomenų rinkinį su sintetiniais duomenimis visi 3 naujai apmokyti modeliai klasifikavo netipinius širdies plakimo garso signalus ženkliai tiksliau. SVM klasifikavo netipinius signalus net 20% tiksliau ir pranašok atsitiktinių miškų algoritmą. Tipinius garso signalus, deja, visi modeliai maždaug 5% procentais klasifikavo prasčiau. Taigi, nors ir tipinių signalų klasifikavimas šiek tiek suprastėjo, netipinių garso signalų klasifikavimas tapo ženkliai tikslesnis ir tai parodo kokią įtaką duomenys turi dirbtinio intelekto klasifikavimo algoritmams ir net sintetiniai duomenys gali turėti stiprų poveikį siekiant atpažinti širdies plakimo garsus tiksliau.

Literatūra

- [1] Cardiovascular diseases (CVDs). URL: <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>
- [2] ALKHODARI, M. - FRAIWAN, L. Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. In *Computer Methods and Programs in Biomedicine*. 2021. Vol. 200, p. 105940. doi: 10.1016/j.cmpb.2021.105940
- [3] CLIFFORD, G.D. ir kt. Recent advances in heart sound analysis. In *Physiological Measurement*. 2017. Vol. 38, no. 8, p. E10. doi: 10.1088/1361-6579/aa7ec8
- [4] KRISHNAN, P.T. ir kt. Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network. In *Physical and Engineering Sciences in Medicine*. 2020. Vol. 43, no. 2, p. 505–515. doi: 10.1007/s13246-020-00851-w
- [5] ZABIHI, M. ir kt. Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. In *2016 Computing in Cardiology Conference (CinC)*. 2016. p. 613–616.
- [6] SHUVO, S.B. ir kt. CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification Using Heart Sound Recordings. In *IEEE Access*. 2021. Vol. 9, p. 36955–36967. doi: 10.1109/ACCESS.2021.3063129
- [7] CHEN, Y. ir kt. Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network. In *Medical & Biological Engineering & Computing*. 2020. Vol. 58, no. 9, p. 2039–2047. doi: 10.1007/s11517-020-02218-5
- [8] DEMIR, F. ir kt. Towards the classification of heart sounds based on convolutional deep neural network. In *Health Information Science and Systems*. 2019. Vol. 7, no. 1, p. 16. doi: 10.1007/s13755-019-0078-0
- [9] CHEN, W. ir kt. Deep Learning Methods for Heart Sounds Classification: A Systematic Review. In *Entropy*. 2021. Vol. 23, no. 6, p. 667. doi: 10.3390/e23060667