

**DETEKCIJA RESPIRATORNIH ANOMALIJA DUBOKIM UČENJEM  
DEEP LEARNING DETECTION OF RESPIRATORY ANOMALIES**Nikola Tomić, *Fakultet tehničkih nauka, Novi Sad***Oblast – ELEKTROTEHNIKA I RAČUNARSTVO**

**Kratak sadržaj** – Respiratorne bolesti izazivaju ogroman zdravstveni, ekonomski i društveni teret i treći su vodeći uzrok smrti širom sveta i značajan teret za javni zdravstveni sistem. Ulažu se značajni istraživački napori posvećeni poboljšanju rane dijagnoze i rutine praćenja pacijenata sa respiratornim oboljenjima, kako bi se omogućile pravovremene intervencije. Respiratorne anomalije koje se mogu prikupiti stetoskopom se klasifikuju kao diskontinualno (pucketanje) ili kontinuirano (piskanje). Takve anomalije u formatu snimljenog zvuka mogu se prikazati i vizuelno u formi spektrograma. Kako su konvolucione neuronske mreže dominantne u polju klasifikacije slika, moguće ih je obučiti i za detekciju respiratornih anomalija. U ovom radu predstavljeni pristup korišćenjem konvolucione neuronske mreže predstavlja značajno poboljšanje u odnosu na prethodne pristupe i dobijeni rezultati su među najboljima nad treniranim skupom podataka.

**Ključne reči:** CNN, ICBHI, Konvolucione neuronske mreže

**Abstract** – Respiratory diseases cause huge health, economic and social burden and are the third leading cause of death worldwide and a significant burden on the public health system. Therefore, significant research efforts have been made to improve the early diagnosis and routine monitoring of patients with respiratory diseases to enable timely interventions. Respiratory anomalies collected with a stethoscope are classified as discontinuous (crackling) or continuous (wheezing). Such anomalies in the format of recorded sound can also be displayed visually in the form of spectrograms. As convolutional neural networks are dominant in image classification, it is possible to train them to detect respiratory anomalies. The approach presented in this paper based on convolutional neural networks represents a significant improvement over previous approaches and the results obtained among the best on the trained data set.

**Keywords:** CNN, ICBHI, Convolutional Neural Networks

**1. UVOD**

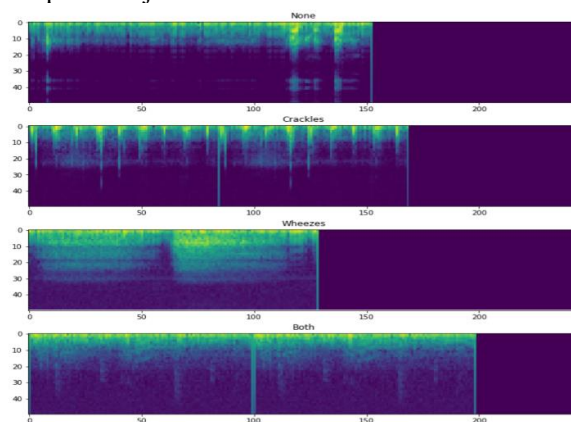
Sprovedena su mnoga istraživanja usredsređena na auskultaciju i karakteristike zvukova respiratornog sistema – kako su direktno povezani sa kretanjem vazduha, prome

**NAPOMENA:**

**Ovaj rad proistekao je iz master rada čiji mentor je bila dr Jelena Slivka, vanr. prof.**

ne unutar plućnog tkiva i položaj sekreta unutar traheo-bronhijalnog stabla. To ih čini vrednim pokazateljem respiratornog zdravlja i respiratornih poremećaja [1].

Dijagnoza i praćenje zasnovano na auskultaciji respiratornih stanja u velikoj meri zavise od prisustva prigodnih zvukova i na izmenjenom prenosu karakteristike zida grudnog koša. Pucketanja su diskontinualne, eksplozivne i nemuzičke slučajnosti koji se često javljaju kod kardiorespiratornih oboljenja [3]. Obično se klasifikuju kao fini i grubi pucketavi po njihovom trajanju, glasnoći, tonu, trenutka pojave u respiratornom ciklusu, odnosa prema kašljanju i promeni pozicije osluskivanja [4]. Zvižduci su prigodni respiratorni zvuci koji obično traju duže više od 250 ms. Uobičajeni su klinički znak kod pacijenata sa opstruktivnim bolestima disajnih puteva kao što su astma i hronična opstruktivna plućna bolest [5]. Moguće klase predikcije i njihov prikaz u formatu spektrograma na mel skali prikazan je na slici 1.



Slika 1. Primer spektrograma zvuka bez anomalija, pucketanja, zvižduka i pucketanja i zvižduka prikazani odozgo nadole u istom redosledu.

Poslednjih godina postignuti su značajni rezultati u obradi slika pomoću neuronskih mreža. Ovo se delimično može pripisati odličnim performansama dubokih konvolucionih neuronskih mreža za detektovanje i transformaciju informacija na apstraktno visokom nivou. Kako se zvuk može prikazati u vidu spektrograma kao dvodimenzionalni prikazi audio frekvencijskih spektara u jedinici vremena, moguće je analizirati i obraditi ga pomoću konvolucionih neuronskih mreža. U toku procesa optimizacije hiperparametara sistema razmatrane su dve različite arhitekture konvolucionih mreža. Arhitektura koja je specijalizovana za klasifikaciju zvuka, odnosno, ljudskog govora, davala je u proseku bolje rezultate u odnosu na

arhitekturu namenjenu za klasifikaciju slika kao sa *ImageNet* [2] takmičenja .

U narednom poglavlju detaljnije je predstavljena postojeća literatura, kao i poređenje sa našim rešenjem. Arhitektura modela i korišćene tehnike opisane su u poglavlju 3, a poglavlje 4 sadrži analizu dobijenih rezultata i mogućih poboljšanja. Poglavlje 5 predstavlja sumarizaciju celog rada.

## 2. PRETHODNA REŠENJA

U radu [6] predstavljen je metod zasnovan na skrivenim Markovljevim modelima (*Hidden Markov Models*, HMM) u kombinaciji sa Gausovim modelima smeše (*Gaussian Mixture Models*, GMM) za klasifikaciju respiratornih zvukova. Sličnosti u pogledu predprocesiranja signala u opisanom radu uključuju smanjenje uzorkovanja na 4 kHz i filtriranje frekvencija *bandpass* filterom kako bi se smanjio uticaj spoljnjih smetnji i šuma. Ulazne karakteristike zvuka su cepstralni koeficijenti mel-frekvencije (*Mel Frequency Cepstral Coefficients*, MFCC) ekstrahovani u opsegu između 50 Hz i 2000 Hz u kombinaciji sa njihovim prvim derivatima. Najbolji rezultat postignut u drugoj fazi evaluacije *ICBHI Challenge* je 39,56. Ovakva evaluacija je direktno uporediva sa pristupom predloženim u ovom radu jer sledi tačno specificirane podele test i trening skupa koje predlažu pravila ICBHI, kao i zvaničnu formulu bodovanja. Rad je osvojio drugo mesto na takmičenju, i predstavlja odličnu referentnu tačku za dalje napredovanje i upoređivanje sa *deep learning* pristupima.

U radu [7] je predložena konvoluciona neuronska mreža za detekciju piskanja nad plućnim zvukovima. Autori tvrde da postiže tačnost od 99% i 0,96 AUC u skupu podataka koji su prikupili. Takođe tvrde da pristup predstavlja trenutni *state-of-the-art* za detekciju piskanja i da je robustan na pomeranja po vremenskoj osi i na spoljašnje smetnje. Korišćeni podaci u radu [7] prikupljeni su iz nekoliko skupova podataka iz više izvora i više od osam puta manji od ICBHI skupa u pogledu broja respiratornih ciklusa. Akcenat u eksperimentisanju gde su se autori posvetili doprinosu augmentacije i uticaja na krajnje performanse. Predloženo rešenje je rezultovalo značajnim poboljšanjima i prednostima u pogledu

robustnosti modela u odnosu na druge statističke modele kao što su vektorske mašine (SVM), K-najbližih suseda (KNN), Logistička regresija (LR), *Gradient Boosting Machine* (GBM), *Random Forests* (RF) itd.

U radu [8] za klasifikaciju respiratornih anomalija (pucketanja i zvižduka), takođe su evaluirani su pristupi zasnovani na konvolucionim neuronskim mrežama. Autori su testirali i ocenili performanse nekoliko arhitektura konvolucionih mreža koje su bile razvijane za potrebe klasifikacije slika nad skupom podataka *ImageNet*, kao što su: *AlexNet*, *VGG16* i *ResNet-50*, gde je *VGG16* arhitektura sa F1 merom od 0.55 značajno poboljšanje u odnosu na prethodne pristupe na ICBHI skupom podataka.

## 3. METOD

U narednim poglavljima izloženi su skup podataka, arhitektura i trening modela.

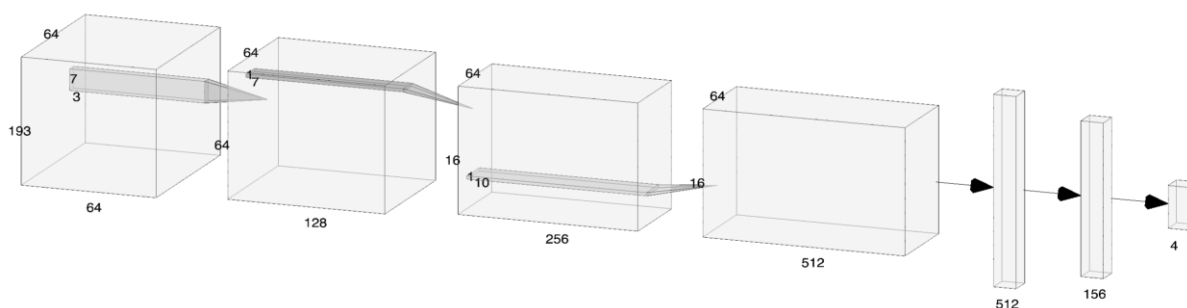
### 3.1. Skup podataka

Konferencija o bio medicinskoj i zdravstvenoj informatici (ICBHI) je prvi naučni izazov organizovan je sa glavnim ciljem razvoja algoritama sposobnih za karakterizaciju snimaka respiratornog zvuka izvedenih iz kliničkog i ne kliničkog okruženja. Bazu podataka kreirala su dva istraživačka tima u Portugalu i Grčkoj, a ona uključuje 920 snimaka pribavljenih od 126 pacijenata. Zabeleženo je ukupno 6898 ciklusa disanja. Baza podataka sastoji se od ukupno 5,5 sati snimaka od kojih 1864 sadrži pucketanje, 886 sadrži zvižduke, a 506 sadrži i pucketanje i piskanje. Svaki zvuk je sniman nad jednom od sedam mogućih pozicija na gornjem delu tela pacijenta. Nivoi buke u nekim ciklusima disanja su visoki, što simulira stvarne životne uslove.

### 3.2. Arhitektura

Arhitektura konvolucione neuronske mreže inspirisana je drugo plasiranom arhitekturom rešenja na *Kaggle's Tensorflow speech recognition competition 2017* [19]. Glavna ideja ove arhitekture bila je stvaranje modela koji bi različito tretirao vreme i frekvenciju. Arhitektura je prikazana na slici 2.

Svaki blok prikazan na slici 2 predstavlja konvolucioni



Slika 2. Arhitektura konvolucione neuronske mreže sa prikazanim dimenzijama svakog sloja i filtera.

sloj praćen *max-pooling* slojem sa nacrtanim dimenzijama filtera. Poslednja tri sloja predstavljaju potpuno povezane slojeve.

Sledi opis svakog sloja iz opisane arhitekture – nazvan imenom klase, sa tačnim parametrima:

1. *Conv2D (64, [7,3])* - Filter šumova i izdvajanje osnovnih obeležja
2. *MaxPool ([1,3])* - Vraćanje na  $N/3$  frekvencijske funkcije, gde  $N$  predstavlja broj mel filtera
3. *Conv2D (128, [1,7])* - Lokalno prepoznavanje šablona po frekventnom opsegu
4. *MaxPool ([1,4])* - Dozvoljava varijacije kod zvuka pacijenta [20]
5. *Conv2D (256 [1,10])* - Izdvajanje funkcija za svaki od frekvencijskih opsega različito i kompresija frekvencijske dimenzije
6. *Conv2D (512, [7,1])* - Traženje povezanih šablona u kratkom vremenskom interval
7. *GlobalMaxPool* - Povezivanje svih komponenti
8. *Dense (256) + Dropout* - Tumačenje obeležja
9. *Dense (4) + Softmax* - Izlazni sloj.

U datom opisu *Conv2D* predstavlja konvolucioni sloj, a *Dense* potpuno povezane slojeve.

### 3.3. Trening

Kompletan sistem (hiperparametri modela i proces augmentacije i predprocesiranja) deo je petlje optimizacije hiperparametara implementirane uz pomoć *Hyperopt python* biblioteke [10]. *Tree of Parzen Estimators* je odabrani algoritam koji se koristi za optimizaciju hiperparametara zbog većeg broja hiperparametara koji generišu široki prostor za pretragu. Funkcija koja se optimizuje je zvanična formula bodovanja *ICBHI 2017* [11]. Svaki eksperiment (skup hiperparametara) je obučavan maksimalno 35 epoha zbog velikog prostora pretrage i ograničenih resursa. Parametri treniranja kao što su stopa učenja (*learning rate*) i *batch size* variraju po vrednostima koje su izabrane empirijskim opsezima a konkretne vrednosti izabrane algoritmom optimizacije.

## 4. REZULTATI I DISKUSIJA

Model sa najboljim performansama dobijen optimizacijom hiperparametara pod nazivom „*iconic-flower-150*“ prikazan u tabeli 1. Tabela 1 takođe predstavlja poređenja dobijenih rezultata sa najboljim radovima nad istim skupom podataka.

Iz kolona *sensitivity* i *specificity* table 1, može se zaključiti da model ima problem sa razlikovanjem zdravih respiratornih ciklusa od ostale tri oznake. Ovaj problem je posebno uočljiv kod piskanja i očekivan je kod dosta ređe klase pucketanja i piskanja, što je potvrđeno matricom

konfuzije.

*ICBHI 2017* je izuzetno izazovan skup podataka sa audio zapisima koji su snimljeni u ne idealnim, realnim uslovima sa dosta smetnji i šuma. Napredne tehnike za procesiranja digitalnog signala u smislu suzbijanja šuma bi sa velikom verovatnoćom bile od koristi u ovakvom okruženju. Ideja za potencijalno poboljšanje otpornosti mreže na šum jeste mogućnost treniranja generativne mreže (*Generative Adversarial Network, GAN*) koja bi proizvodila dinamički šum i ubacivala u snimke. Velika razlika u pacijentima (godine, veličina, pol, itd.) i različita oprema kojom su zapisi snimani čini ovaj skup podataka još izazovnijim. U ovakvom slučaju veća količina podataka bi svakako bila značajna u pogledu poboljšanja performansi *deep learning* modela, što se pokazalo i u eksperimentima sa agresivnijom augmentacijom retkih klasa. Korišćenjem biblioteke za optimizaciju hiperparametara uz veći broj eksperimenata otkriveni su značajniji parametri u celom sistemu, međutim, treba napomenuti da su se zbog vremenskih i računarskih ograničenja svi modeli trenirali maksimalno 35 epoha. U radu [8] model sa kojim su rezultati poređeni u tabeli 1 bio je treniran 150 epoha nad istim *ICBHI* skupom podataka. Imajući u vidu da su trenirani modeli i dalje konvergirali i funkcija greške se smanjivala, dužim treniranjem bi se potencijalno dobili još bolji rezultati.

## 5. ZAKLJUČAK

U ovom radu je predložen pristup dubokog učenja pri detekciji i klasifikovanju respiratornih anomalija. Respiratorne bolesti izazivaju ogroman zdravstveni, ekonomski i društveni teret, treći su vodeći uzrok smrti širom sveta i značajan teret za globalno zdravstveni sistem.

Poslednjih godina postignuti su značajni rezultati u obradi slika pomoću konvolucionih neuronskih mreža. Ovo se delimično može pripisati odličnim performansama dubokih konvolucionih neuronskih mreža za detektovanje i transformaciju informacija na apstraktnom nivou kod slika. Kako se zvuk može prikazati u vidu spektograma kao dvodimenzionalni prikazi audio frekvencijskih spektara u jedinici vremena, moguće je analizirati i obraditi ga pomoću konvolucionih neuronskih mreža. U ovom radu, kao i u relevantnoj literaturi, obavljena su poređenja i opisane prednosti *deep learning* pristupa ovom problemu u odnosu na tradicionalne statističke modele sa ručnim odabirom obeležja. Rezultati dobijeni u ovom radu trenutno predstavljaju najbolje rangirano rešenje po pravilima *ICBHI 2017* takmičenja. Rezultati su takođe upoređeni sa sličnim *deep learning* pristupom nad istim skupom podataka. Međutim, treba napomenuti da se u pomenutom radu nije koristio zvanični test skup za evaluaciju, sto može dovesti do optimističnih rezultata

Tabela 1. Poređenje dobijenih rezultata u ovom radu (poslednji red) sa rezultatima iz radova [6] i [8].

	Sensitivity	Specificity	Precision	F1 score	ICBHI score
Demir et al.	0.53	0.83	0.57	0.55	/
Jakovljević et al.	0.41	0.52	/	/	36.98
<i>iconic-flower-150</i>	<b>0.56</b>	<b>0.55</b>	<b>0.54</b>	<b>0.54</b>	<b>53.4</b>

kao što je napomenuto u radu [6]. Arhitektura koja je predložena u ovom rešenju inspirisana je konvolucionom neuronskom mrežom dizajniranom za prepoznavanje ljudskog govora. Eksperimentima u ovom radu pokazano je da takva specifična arhitektura ima prednosti u odnosu na *AlexNet* arhitekturu koja je dizajnirana pre svega za klasifikaciju slika. Ova prednost može se pripisati pažljivo osmišljenim slojevima i dimenzijama filtera koji različito tretiraju vremensku i frekvencijsku osu.

#### Kratka biografija:



**Nikola Tomić** rođen je 1996. godine u Smederevu. Osnovne akademske studije završio je 2019. godine na Fakultetu tehničkih nauka, na kom brani i master rad 2021. godine iz oblasti Elektrotehnike i računarstva – Softversko inženjerstvo i informacione tehnologije.

kontakt: [tomic.nikola.uns@gmail.com](mailto:tomic.nikola.uns@gmail.com)

#### 6. LITERATURA

- [1] Gibson GJ, Loddenkemper R, Lundbäck B, Sibille Y. Respiratory health and disease in Europe: the new European Lung White Book. *Eur Respir J.* 2013;42:559–63.
- [2] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge." *International journal of computer vision* 115.3 (2015): 211-252.
- [3] Piirila P, Sovijarvi AR. Crackles: recording, analysis and clinical significance. *Eur Respir J. Eur Respiratory Soc*; 1995;8(12):2139–48.
- [4] Sarkar M, Madabhavi I, Niranjana N, Dogra M. Auscultation of the respiratory system. *Ann Thorac Med. Medknow Publications*; 2015;10(3):158.
- [5] Sovijarvi ARA. Characteristics of breath sounds and adventitious respiratory sounds. *Eur Respir Rev.* 2000;10:591–6.
- [6] Jakovljević, Nikša, and Tatjana Lončar-Turukalo. "Hidden markov model based respiratory sound classification." *International Conference on Biomedical and Health Informatics*. Springer, Singapore, 2017.
- [7] Kochetov, Kirill, et al. "Wheeze detection using convolutional neural networks." *EPIA Conference on Artificial Intelligence*. Springer, Cham, 2017.
- [8] Demir, Fatih, Abdulkadir Sengur, and Varun Bajaj. "Convolutional neural networks based efficient approach for classification of lung diseases." *Health information science and systems* 8.1 (2020): 1-8.
- [9] <https://www.kaggle.com/c/tensorflow-speech-recognition-challenge>
- [10] <https://hyperopt.github.io/hyperopt/>
- [11] <https://bhichallenge.med.auth.gr/>