



DETEKCIJA COVID-19 SLUČAJEVA SA RTG SNIMAKA UPOTREBOM DUBOKIH NEURONSKIH MREŽA

COVID-19 CASES DETECTION FROM X-RAY IMAGES BY USING DEEP NEURAL NETWORKS

Stefan Orčić, *Fakultet tehničkih nauka, Novi Sad*

Oblast – INŽENJERSTVO INFORMACIONIH SISTEMA

Kratka sadržaj – *Pandemija COVID-19 ima razarajući efekat na zdravlje stanovništva širom sveta. Zbog toga je presudni korak, u borbi protiv ove pandemije, kontrola širenja bolesti. Kapacitet i kvalitet laboratorijskog ispitivanja je izazovni zadatak, pa alternativne metode ispitivanja igraju značajnu ulogu u ovoj borbi. Na osnovu toga, jedan od novih pristupa oslanja se na analizu radiološkog snimanja pomoću radiografije grudnog koša za dijagnozu, procenu i proveru stadijuma infekcije COVID-19. Razvoj automatizovanog alata koji će koristiti veliki broj rendgenskih zraka za klasifikaciju bio bi od velike važnosti kada bi se obuhvatio veliki broj slučajeva. Prethodnih godina, savremene arhitekture revolucionarnih konvolutivnih neuronskih mreža pokazale su izvanredne rezultate u brojnim zadacima medicinske klasifikacije. Motivisani ovim, eksperimenti izvedeni u ovom radu analiziraju njihovu upotrebu u zadatku otkrivanja slučajeva COVID-19 klasifikacijom rendgenskih fotografija, uz primenu strategije prenosnog učenja (engl. transfer learning) na unapred obučenoj ImageNet mreži, uz upotrebu različitih tehnika pred-procesiranja (engl. pre-processing). Rezultati predstavljeni u ovom radu zaključuju da bi duboke konvolucione neuronske mreže mogle izvući radiološke vizuelne karakteristike koje su u korelaciji sa biomarkerima koji su povezani u slučajevima COVID-19 sa velikom tačnošću.*

Ključne reči: *klasifikacija COVID-19, konvolucione neuronske mreže, duboko učenje.*

Abstract – *The COVID-19 pandemic has a destructive effect on health on the population worldwide. The crucial step in the fight against it is to control the spread of the disease while screening the broad number of suspected cases for appropriate quarantine and treatment as priority. Capacity and quality of laboratory testing is a challenging task so alternative methods in testing plays a significant role in this fight. Based on that, one of the novel approaches relies on analyzing radiological imaging using chest radiography for diagnosis, assessment and staging of COVID-19 infection. Developing an automated tool that will utilize the large number of X-rays for classification would be of great importance when covering a large number of cases. In the*

previous years, the state-of-the-art convolutional neural networks architectures showed outstanding results in numerous medical classification tasks. Motivated by this, experiments conducted in this paper are analyzing their usage in the task of detection of COVID-19 cases by X-ray image classification along with employing the transfer learning strategy on pre-trained grayscale ImageNet while also using various pre-processing techniques. The results presented in this paper come to the conclusion that deep convolutional neural networks could extract radiological visual features that correlates with biomarkers that are related to COVID-19 cases with the high accuracy.

Keywords: *COVID-19 classification, convolutional neural networks, deep learning.*

1. UVOD

Prva tema u svetu u poslednjih dve godine jeste bolest koronavirus (COVID-19). U decembru 2019. pojavio se na pijaci morskih plodova Huanan u gradu Vuhan, provincija Hubei, Kina, nakon što se brzo proširio po celom svetu. Za samo dva meseca virus se proširio iz Vuhana u čitavu Kinu i još 33 zemlje. Svetska zdravstvena organizacija proglasila je pandemiju 11. marta 2020. godine. Ovaj virus može da zarazi ljude svih starosnih grupa. COVID-19 je izuzetno zarazan i vrlo brzo se proširio Globalno. Tipične kliničke karakteristike COVID-19 uključuju groznicu, kašalj, upalu grla, glavobolju, umor, bolove u mišićima i otežano disanje. Najvažnija stvar u bilo kojoj pandemiji je brzo otkrivanje zaraženih, lečenje zaraženih i otkrivanje svih osoba sa kojima je zaražena osoba imala kontakt, jer je period inkubacije približno 5 dana [1, 2].

Budući da COVID-19 brzo napreduje, neophodno je otkriti bolest u ranoj fazi. Laboratorijsko otkrivanje virusne RNK pomoću qRT-PCR-a (engl. Quantitative Reverse Transcription Polymerase Chain Reaction - qRT-PCR) u respiratornim uzrocima, za koje se sumnja da sadrže virus COVID-19, se smatra jednim od bitnih kriterijuma za dijagnozu COVID-19 bolesti.

Takođe, standardni CT grudnog može se može koristiti za dijagnozu COVID-19. Kao što je nedavno objavljeno, CT grudnog koša pokazuje tipične karakteristike kod gotovo svih pacijenata sa COVID-19 [3]. Klasifikacija na osnovu CT grudnog koša uključuje stručnjaka za radiologiju i znatno vreme.

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio prof. dr Srđan Sladojević.

Automatska analiza je poželjna da bi se uštedelo vreme medicinskog radnika. Za automatsko otkrivanje COVID-19, važno je da postoje medicinski snimci, grudnog koša, ljudi obolelih od bolesti COVID-19 i ljudi koji nisu oboleli od COVID-19 bolesti. Na osnovu činjenice da su mašinsko učenje i duboko učenje postali ustaljene discipline u obrascima analize na osnovu podataka, obe oblasti su takođe našle primenu u detekciji COVID-19. Širok spektar zadataka računarske vizije mogao bi da koristi konvolucione neuronske mreže (CNN) kao izuzetno moćne alate. Činjenica koja dokazuje ovu izjavu je da duboki CNN-ovi automatski uče apstrakcije srednjeg i visokog nivoa dobijene iz neobrađenih podataka. Zanimljivo je da se u medicinskom snimanju moglo videti da tačna dijagnoza bolesti zavisi od sticanja i interpretacije fotografije [4, 5]. Ključni faktori za poboljšanje dijagnoze su kompjuterizovani alati koji se odnose na analizu fotografije i mašinsko učenje. To se postiže olakšavanjem identifikovanja nalaza koji zahtevaju lečenje i podržavanjem stručnog procesa rada. Među ovim alatima, duboko učenje se brzo pokazuje kao najsavremenija osnova, što dovodi do poboljšane tačnosti. Takođe duboko učenje otvara nove granice u analizi podataka sa stopama napretka koji se prethodno nisu iskusili [6].

2. EKSPERIMENT

U ovoj sekciji je detaljnije objašnjena metodologija rada, skup podataka, model i rezultati eksperimenta

2.1. Metodologija rada

Eksperiment korišćen za ovaj rad podeljen je u nekoliko faza koje uključuju pripremu skupa podataka zajedno sa potrebnim pred-procesiranjem i augmentacijom, dizajniranjem arhitekture duboke konvolucione neuronske mreže zajedno sa dobro strukturisanim fazama treninga, testa i validacije.

S obzirom na to da do trenutka pisanja rada nema većih skupova podataka COVID-19, potrebno je ručno prikupljanje i kombinovanje trenutno javno dostupnih skupova podataka. Autori [7] pružili su detaljan pregled postojećih skupova podataka CXR (rendgenske fotografije grudnog koša) koji bi mogli da se koriste za dalja istraživanja.

2.2. Skup podataka

Rtg snimci u skupu podataka, nazvanom COVID-ChestXRy, su izvučeni iz veb lokacija, online publikacija ili direktno iz PDF-a. Prilikom prikupljanja očuvan je kvalitet samih rtg snimaka.

Snimci predstavljaju radiološke nalaze i labelirani su od strane radiologa. U vreme pisanja ovog dokumenta sastoji se od 434 fotografije slučajeva COVID19. Skup podataka COVID-ChestXRy2017 sadrži 5857 rendgenskih fotografija prikupljenih u

Medicinskom centru za žene i decu u Guangdžou, Kina. Te fotografije su prikupljene u tri različite klase:

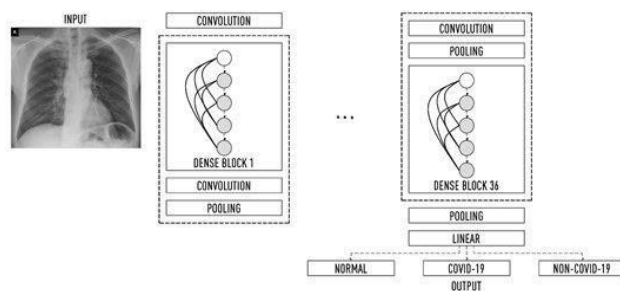
- normalni,
- bakterijski i
- virusni slučajevi [8].

Da bi se prevazišla moguća prezasićenost (engl. Overfitting) i zbog činjenice da prikupljeni skup podataka nije uravnotežen, nad fotografijama sa potvrđenom COVID-19 dijagnozom je izvršena augmentacija. Metode augmentacije koje su korišćene u ovom eksperimentu obuhvatale su rotaciju koja je sačuvala veličinu (slučajna rotacija između 10-15 stepeni) i smicanje (po x-osi ili y-osi za 10 stepeni). Takođe su korišćene i druge transformacije u perspektivi poput iskrivljavanja i nagnjanja. Sve fotografije korišćene za ovaj eksperiment skalirane su na 224x224 piksela.

Celokupna augmentacija izvršena je korišćenjem Augmentor alata, koji je specijalizovan za biomedicinsku augmentaciju fotografija, koji su razvili autori rada [9]. Augmentacija i promena veličine vršeni su u fazi pred-procesiranja, pre formiranja mreže i treniranja modela, kako bi se ubrzale faze treninga. Nakon faze pred-procesiranja fotografija, klasa COVID-19 je povećana na 2170 fotografija.

2.3. Model

Arhitektura neuronske mreže, ResNetCOVID19, koja je korišćena u ovom radu, sastoji se od 36 dense bloka, gde dva susedna bloka služe kao prelazni slojevi, koji menjaju veličine mapa karakteristika pomoću konvolucije i udruživanja, a koji su prikazani na slici 19. Ulazni sloj prihvata rendgenski snimak 224x224 piksela, dok izlazni sloj predstavlja finalnu klasifikacija se deli u tri klase: Normalna, Covid-19 i Ne-Covid19 pluća.



Slika 1 - Prikaz ResNetCOVID19 mrežne arhitekture

Skup podataka korišćen u eksperimentalnoj fazi ovog rada podeljen je na 80-20, koji predstavlja odnos validacije i treninga. Podaci o treningu su podeljeni na 70-30, odnos trening-test.

Predložena duboka arhitektura iz prethodnog odeljka prvo je obučena na ImageNet-u nakon pretvaranja tih fotografija u sive tonove koristeći uobičajenu transformaciju. Nakon te faze, razvijen je model klasifikatora COVID-19 primenom strategije prenosnog učenja tokom faze treninga na skupu podataka.

SGD optimizator sa zakazanom stopom učenja primenjen je tokom obe faze treninga, na sivim tonovima ImageNet i COVID-19 setu podataka.

2.4. Rezultati

Iz razloga što je tema nova, u trenutku pisanja ovog rada nema mnogo akademskih radova i predloženih pristupa koji se bave zadatkom klasifikacije COVID-19 zasnovanog na rendgenskim snimcima. Zbog toga je

COVID-Net model trenutno osnovni model za metrike performansi. Autori rada [10] dozvolili su pristup otvorenom kodu unapred obučeni modela [11] kako bi se mogli koristiti kao referentni model za dalja poboljšanja. Za eksperiment, kao referentna vrednost korišćen je najnoviji model COVID-Net, COVID-Net-CXR4.

Covid-Net model ima 92.4 postotak tačnosti na test skupu podataka, dok mu je postotak osetljivosti u odnosu na klase: 92% za normalna pluća, 90.1% za pluća koja nemaju nikakvo oboljenje i 96.9% za pluća inficirana virusom COVID19.

Tabela 1 - Prikaz tačnosti ResNetCOVID19

Arhitektura	Tačnost (%)
ResNetCOVID19	94.1

Tabela 2 - Prikaz osetljivosti ResNetCOVID19

Arhitektura	Normalna	Ne-COV19	COV19
ResNetCOVID19	91.4	92.3	97.3

Na osnovu rezultata prikazanih u Tabeli 1, može se zaključiti da arhitektura postiže visoku tačnost na test skupu. Na osnovu tabele 2, može se uočiti da je u ResNetCOVID19 modelu visok procenat osetljivosti, koja je najvažniji pokazatelj, kao i broj propuštenih slučajeva COVID-19 treba svesti na minimum.

3. ZAKLJUČAK

Ovaj rad je predložio novu arhitekturu, ResNetCOVID19, koja je prethodno obučena na fotografijama u sivim tonovima ImageNet modela, zajedno sa odgovarajućom augmentacijom rendgenskih fotografija COVID-19 što je poboljšalo tačnost i osetljivost trenutno predloženih rešenja na javno dostupnim skupovima podataka i sprečilo prezasićenje (engl. Overfitting).

Jedan od trenutnih izazova u ovom klasifikacionom zadatku jeste mala količina javno dostupnih podataka, što donosi mnogo prepreka za uspeh obuke dubokog modela. Izdvajanje COVID-19 karakteristika i način rukovanja subjektivnim karakteristikama u skupovima podataka predstavljaju najzahtevnije zadatke u ovoj oblasti istraživanja i potrebno je pažljivo rukovati tim podacima.

Obećavajući rezultati nedavnih istraživanja i razni projekti otvorenog koda bi potencijalno mogli da ubrzaju ovu novouvedenu temu istraživanja.

Dalji pravci uključuju primenu različitih tehnika vizuelizacije koje bi mogle doneti više razumevanje specifičnih karakteristika COVID-19. Takođe, moguća su i dalja istraživanja koja uključuju kombinovanje klasifikacije rendgenskih fotografija zajedno sa ostalim podacima o pacijentu.

4. LITERATURA

- [1] Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9), 577-582.
- [2] Mehta, P., McAuley, D. F., Brown, M., Sanchez, E., Tattersall, R. S., Manson, J. J., & HLH Across Speciality Collaboration. (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet*, 395(10229), 1033.
- [3] Chung, M., Bernheim, A., Mei, X., Zhang, N., Huang, M., Zeng, X., ... & Jacobi, A. (2020). CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology*, 295(1), 202-207.
- [4] Bar, Y., Diamant, I., Wolf, L., & Greenspan, H. (2015, March). Deep learning with non-medical training used for chest pathology identification. In *Medical Imaging 2015: Computer-Aided Diagnosis* (Vol. 9414, p. 94140V). International Society for Optics and Photonics.
- [5] Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., & Greenspan, H. (2015, April). Chest pathology detection using deep learning with non-medical training. In *2015 IEEE 12th international symposium on biomedical imaging (ISBI)* (pp. 294-297). IEEE.
- [6] Greenspan, H., Van Ginneken, B., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5), 1153-1159.
- [7] Tartaglione, E., Barbano, C. A., Berzovini, C., Calandri, M., & Grangetto, M. (2020). Unveiling COVID-19 from Chest X-ray with deep learning: a hurdles race with small data. *arXiv preprint arXiv:2004.05405*.
- [8] ChestXRy2017, <https://data.mendeley.com/datasets/rschjbr9sj/2/files/f12eaf6d6023-432f-acc9-80c9d7393433> (pristupljeno u julu 2020.)
- [9] Bloice, M. D., Roth, P. M., & Holzinger, A. (2019). Biomedical image augmentation using Augmentor. *Bioinformatics*, 35(21), 4522-4524.
- [10] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep Residual Learning for Image Recognition.
- [11] COVID-Net Open Source Initiative, <https://github.com/lindawangg/COVID-Net> (pristupljeno u julu 2020.)

Kratka biografija:



Stefan Orčić rođen je u Novom Sadu 1996. god. Diplomski rad na Fakultetu tehničkih nauka iz oblasti Informacionih tehnologija odbranio je 2019. god.
kontakt: orcic.stefan@gmail.com