

ОИМЕ И ПРЕЗИМЕ: Моника Коруноска БРОЈ НА ИНДЕКС: 171062

1. **(20 поени)** За ова прашање ќе треба да најдете оригинален истражувачки труд на сајтот:

[Scholar.google.com](https://scholar.google.com)

Трудот треба да има секција за методи (најчесто поднаслов Methods или Methodology) и да има јасна хипотеза. Бидејќи голем дел од трудовите се достапни само со плаќање (paywalled), на час ви кажавме како да пристапите до нив бесплатно. Целиот колоквиум е поврзан со истиот труд, така што посветете доволно време во изборот на трудот за да можете полесно да ги одговорите сите прашања и задачи.

Цитирајте го избраниот труд користејќи го IEEE стилот на цитирање!

**ОДГОВОР:**

Linda Wang, Zhong Qiu Lin, and Alexander Wong.

"COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images." *Scientific Reports* (2020)

```
@Article{Wang2020,
  author={Wang, Linda and Lin, Zhong Qiu and Wong, Alexander},
  title={COVID-Net: a tailored deep convolutional neural network design for
detection of COVID-19 cases from chest X-ray images},
  journal={Scientific Reports},
  year={2020},
  month={Nov},
  day={11},
  volume={10},
  number={1},
  pages={19549},
  issn={2045-2322},
  doi={10.1038/s41598-020-76550-z},
  url={https://doi.org/10.1038/s41598-020-76550-z}
}
```

2. **(70 поени)** Дизајнирајте научен експеримент што би дал одговор на едно прашање што ве интересира. Експериментот треба да е квантитативен, но не мора да биде реалистичен или лесно изводлив. Опишете ја методологијата на истражувањето во следните категории:

а) Која е хипотезата што ја тестирате? (Хипотезата треба да е потврдна реченица)

Пациентите инфицирани со Covid-19 имаат аномалии во сликите на радиографијата на градниот кош, кои со дизајн на длабока конволутивна нервна мрежа може да се прилагоди за откривање на случаи на Covid-19 од рендгенски слики на градите (CXR).

б) Кои величини/квантитети ќе ги измерите како дел од експериментот?

Датасетот CovidX кој се користи за тренирање и евалуација на моделот на COVID-Net, кој содржи 13,975 CXR слики, од 13,870 пациенти.

Различните случаи што ги содржи CovidX:

- Пациенти кои имаат пневмонија, но не Covid19 и пациенти со Covid19
- Пациенти кои немаат пневмонија и немаат Covid19 и пациенти кои немаат Covid19, но имаат пневмонија

Точноста на тестот.

Чувствителноста на тестот.

Позитивната предиктивна вредност (PPV)

в) Кој статистички метод ќе го користите за да ја тестирате хипотезата?

За да се оцени ефикасноста на предложената хипотеза на COVID-Net, се извршуваат и квантитативна и квалитативна анализа за подобро да се разберат перформансите на детекција и однесувањето на донесувањето на одлуки.

Квантитативна анализа

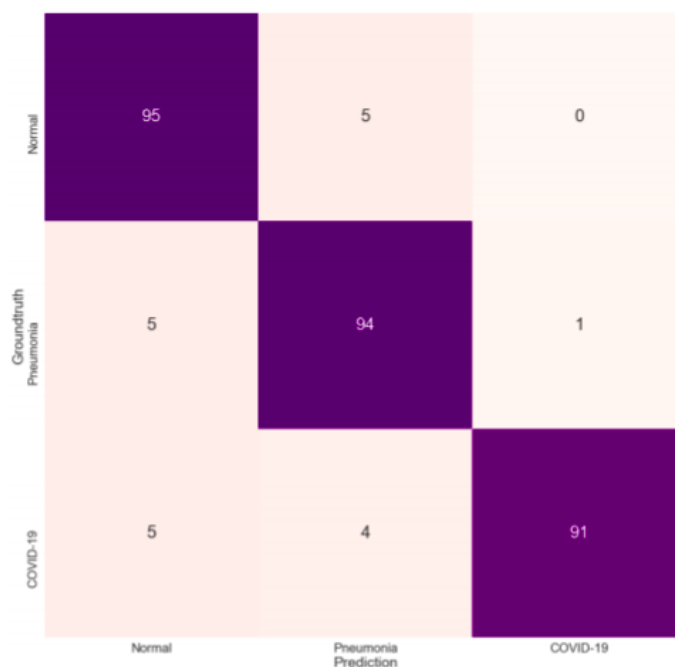
За да ја испитае предложената COVID-Net на квантитативен начин, може да се пресмета точноста на тестот, како и чувствителноста и позитивна предиктивна вредност (PPV) за секој тип на инфекција, на базата на податоци COVIDx.

Потоа, може да се направи подлабоко истражување на тековните ограничувања на предложениот COVID-Net со проучување на чувствителноста и PPV за секој тип на инфекција.

## Квалитативна анализа

Може да се изврши ревизија на предложениот COVID-Net за да се добие подобар увид во тоа како COVID-Net донесува одлуки и да се провери дали детектира и донесува одлуки врз основа на релевантни информации, наместо погрешни информации врз кои се донесуваат одлуки на нерелевантни визуелни индикатори.

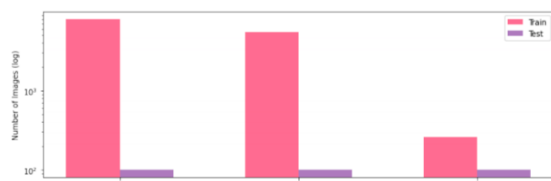
г) Како ќе ги визуелизирате мерењата од б)? Предложете скица во која ќе бидат претставени величините од мерењата. Скицата можете да ја пратите во прилог како дигитална фотографија.



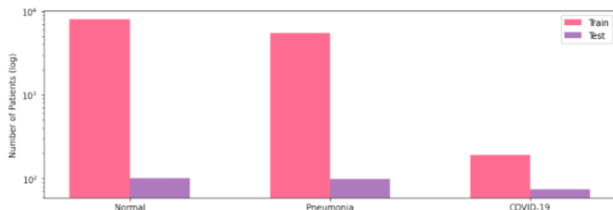
**Figure 6.** Confusion matrix for COVID-Net on the COVIDx test dataset.

**Table 3.** Positive predictive value (PPV) for each infection type. Best results highlighted in **bold**.

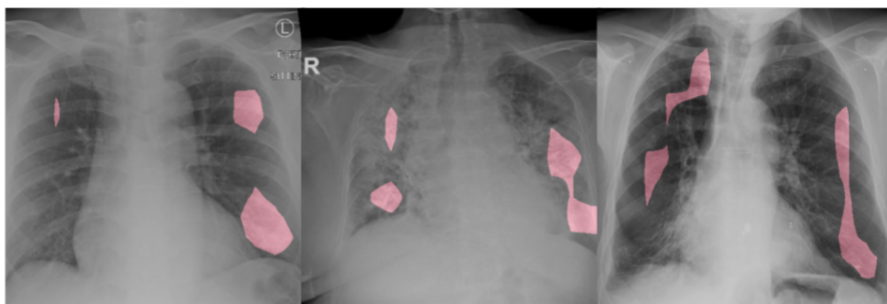
Positive Predictive Value (%)			
Architecture	Normal	Non-COVID19	COVID-19
VGG-19	83.1	75.0	98.4
ResNet-50	88.2	86.8	98.8
COVID-Net	<b>90.5</b>	<b>91.3</b>	<b>98.9</b>



**Figure 3.** CXR images distribution for each infection type of the COVIDx dataset (normal means no infection). (Left bar) number of training images, (right bar) number of test images.



**Figure 4.** Number of patient cases for each infection type of the COVIDx dataset (normal means no infection). (Left bar) number of patient cases for training, (right bar) number of patient cases for testing.



**Figure 7.** Example CXR images of COVID-19 cases from several different patients and their associated critical factors (highlighted in red) as identified by GSInquire<sup>37</sup>.

3. **(120 поени)** Направете Jupyter тетратката поврзана со трудот кој го цитиравте во првото прашање и прикачете ја на GitHub (доколку немате профил креирајте го, ќе ви треба). Линкот од вашиот Github геро мора да биде испратен до 23.59 часот на 14 јуни (сите промени по овој краен рок нема да бидат прифатени). Исто така нема да прифаќаме тетратки хостирани на било кое друго место освен на Github.

а) Тетратката треба да започне со краток опис на трудот (напишан во Markdown). Краткиот опис треба во стотина зборови да објасни зошто е овој труд значаен.

б) Остатокот од тетратката го оставаме на вас. Не заборавате дека испитите ќе

бидат рангирани, така што тие кои ќе имаат најквалитетна тетратка ќе добијат најмногу поени. За да биде кандидат за максимална оценка, тетратката треба да содржи повеќето од овие карактеристики:

- Формули од избраниот труд напишани во LaTeX
- Ќелии со код од избраниот труд кои може да се егзекутираат (полесно е ова да се направи доколку податоците и кодот од трудот се јавно достапни) - Интерактивна визуелизација (Plotly, ipywidgets или други алатки)
- Вметнато лого на журналот во кој е објавен трудот
- Ембедиран мултимедијален запис поврзан со трудот (YouTube видео, podcast, ...)

Целта на ова прашање е да бидете креативни. Понудете ни тетратка која го надополнува оригиналниот PDF и го прави истражувањето да биде покорисно. Доколку трудот ги споделува податоците, тогаш можете да направите и сосема нова визуелизација. Изненадете нè!

4. **(40 поени)** Обидете се да ја подобрите репродуцибилноста на тетратката од прашање 3, со некоја од една од следните три алатки:

- [Binder](#)
- [Docker](#)
- [Google Colab](#)

Идејата е сите фигури и пресметки да можат да се извршат во некоја од наведените алатки. Притоа, прашањето носи исто поени независно од околината (некој што има само Binder ќе биде исто оценет и доколку има Binder со Docker).

P.S. Вашите одговори на испитот треба да бидат прикачени на GitHub (во PDF или друг електронски формат) заедно со Jupyter тетратката и линот да го прикачите.

GitHub: <https://github.com/KKMonika/INSOK>

Binder: <https://mybinder.org/v2/gh/KKMonika/INSOK/master?filepath=Research%20paper.ipynb>