Műszaki Tudományos Közlemények vol. 20. (2024) 74–77.
DOI: Magyar: https://doi.org/10.33895/mtk-2024.20.14
Angol: https://doi.org/10.33894/mtk-2024.20.14

# AI-ALAPÚ RÖNTGENKÉPELEMZÉSI ELJÁRÁS VALÓS IDEJŰ OUT-OF-DISTRIBUTION-DETEKTÁLÁS SEGÍTSÉGÉVEL

# DEVELOPMENT OF ROBUST AI BASED X-RAY ANALYSIS WORKFLOW BASED ON REAL-TIME OUT OF DISTRIBUTION DETECTION

Szabó Lóránt,[1] Weltsch Zoltán[2]

[1] *AI Research Center (Neumann János Egyetem), Kecskemét, Magyarország. szabo.lorant@nje.hu*

[2] *Széchenyi István Egyetem, Győr, Magyarország. weltsch.zoltan@sze.hu*

**Abstract**

In medicine, the role of artificial intelligence (AI) is becoming increasingly common. An important example here is the application of AI in X-Ray analysis, as a known aspect of medical imaging and finding-detection. However, the effectiveness of AI image analysis may be challenging due to the out-of-distribution (OOD) records, i.e. data that significantly differ from the data set used to train the model. These OOD data may result from symptoms, that the model is not prepared for, or even from unpredictable tool behaviour, environmental changes or new errors that have not occurred during the data-gathering phase. This paper shows that with proper OOD analysis the AI-based tool may be prepared for handling “unknown” input data.

**Keywords***: AI, OOD, ID, t-SNE.*

**Összefoglalás**

Az orvostudományban a mesterséges intelligencia (AI) szerepe egyre nagyobb. Kiváló példa erre az AI alkalmazása a röntgenképek elemzésére, ami az orvosi képalkotás során történő képletfelismerés egyik ismert aspektusa. A mesterséges intelligencián alapuló képelemzés alkalmazása szempontjából jelentős kihívással bírnak az úgynevezett out-of-distribution- (OOD-)esetek, melyek a modell betanításához használt adathalmaztól jelentősen különböző adatokat takarnak. OOD-adatok származhatnak olyan tünetekből, amelyekre a modell nincs felkészítve, avagy adódhatnak akár az eszköz előre nem látható viselkedéséből vagy környezeti anomáliákból, esetleg olyan hibákból, amelyek az elsődleges adatgyűjtési fázis során nem fordultak elő. Ez a cikk bemutatja, hogy megfelelő OOD-elemzéssel az AI-alapú eszköz felkészülhet az "ismeretlen" bemeneti adatok kezelésére.

**Kulcsszavak***: AI, OOD, ID, t-SNE.*

## 1. Bevezetés

### 1.1. AI az orvosi képalkotásban

Egyre több kórház használja a különböző mesterségesintelligencia-alapú módszereket az egyes osztályokon. Ezzel együtt az AI használatának egyik legnagyobb kihívása továbbra is az, hogy miként szavatolható annak megfelelő működése. A mesterséges intelligencia hatékony felhasználása érdekében olyan munkafolyamatot kell meghatározni, amelyben a mesterséges intelligencia elsősorban az orvosi felvételeken található elváltozások megtalálására szolgál, míg az orvos ezen iránymutatások alapján diagnosztizál. A legtöbb esetben nem az a probléma, hogy a mesterséges intelligencia esetleg nem képes osztályozni vagy felismerni egy rendellenességet. A probléma akkor merül föl, amikor nem biztos, hogy az adott AI-alapú modell fel van készítve egy adott a beme-

net feldolgozására, és ennek ellenére annak kiértékelését várjuk tőle. Ebben a tanulmányban egy kísérletet mutatunk be mellkas-röntgenfelvételek segítségével, melyek négy osztályba sorolhatók: Normal, Lung Opacity, Pneumonia és Covid. Itt a Covidot új típusú tünetegyüttesnek tekintjük; e tanulmány ismerteti, hogyan kezelhetők azok az adatok, amelyek „ismeretlenek" az éppen használt AI-eszköz számára.

### 1.2. Out-of-distribution-adatok

A robusztus, szabályozott rendszerek esetén különösen fontos a modellek új adatokhoz, valamint az előre nem jelezhető körülményekhez való alkalmazkodóképessége. A mesterségesintelligencia-alapú technikák hatékonyságának növelése terén különösen nagy kihívást jelentenek az OOD-adatok [1]. Az OOD-adatok és ezek eloszlása valamilyen módon eltér a tanító halmaztól (in-distribution- vagy ID-adatok) akár reprezentáció, akár kontextus, akár más tekintetében. Másképpen, ha $\boldsymbol{X}$ és $\boldsymbol{Z}$ adott eloszlású adathalmazok, és egy modellt az $\boldsymbol{X}$-ben lévő $x_1$, $x_2$, ..., $x_n$ adatokra optimalizáltunk, akkor a $\boldsymbol{Z}$-beli $z$ adat akkor és csak akkor OOD, ha

$$\boldsymbol{X} \neq \boldsymbol{Z}$$

Az OOD-adatok jelentősen befolyásolhatják a mélytanulási modellek teljesítményét, azaz gyakran alacsony pontossághoz vagy váratlan viselkedéshez vezetnek. A probléma abból a feltételezésből ered, hogy a tanító adatok minden lehetséges szempontból reprezentatívak; ez a feltételezés azonban a valós esetekben gyakran nem állja meg a helyét.

E tanulmány a hivatkozott módszerekkel egy online adathalmazból [2, 3] elérhető röntgenfelvételek elemzését mutatja be. Ezek a felvételek mint képek szolgálnak elsődleges bemeneti adatként; a diagnózis alapján négy különböző osztályba sorolhatók: Normal, Lung Opacity, Pneumonia és Covid.

## 2. Kiértékelési módszerek

E fejezet a kísérlet módszertani részleteit ismerteti: a megfelelő neurális háló finomhangolása után az OOD-adatok klasszifikációval egy időben történő detektálási lehetőségét mutatja be.

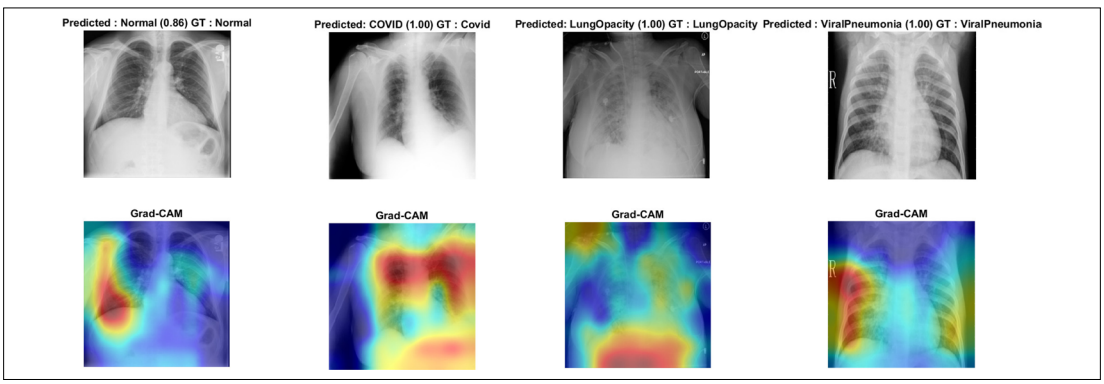
### 2.1. Adatok előkészítése OOD-analízishez

A röntgenfelvételek a korábban említetteknek megfelelően az alábbi osztályokba sorolhatók: Normal, Lung Opacity, Pneumonia és Covid. A négy osztályból álló adathalmaz segítségével a következő stratégiát alkalmazzuk: ahelyett, hogy mind a négy osztály felhasználásával tanítanánk egy osztályozó neurális hálót, egy olyan hálót tanítunk, amely csak az első három osztály, azaz a Normal, a Lung Opacity és a Pneumonia klasszifikációjára képes.

A negyedik osztályt, a Covidot kizárjuk a tanító halmazból, így ez OOD-mintául szolgál. Ez a megközelítés lehetővé teszi a háló teljesítményének tanulmányozását, egyúttal az OOD-adatok által támasztott elvárások tesztelését.

### 2.2. OOD-detekciós algoritmusok

Az OOD-adatok kezelése kritikus fontosságú az AI-alapú alkalmazásokban; nemcsak a modellek teljesítményének javítása szempontjából, hanem a dinamikusan változó környezetben való robusztus viselkedés biztosítása céljából is. Az OOD-detektálás célja az olyan bemeneti adatok azonosítása, amelyek jelentősen eltérnek a tanító adathalmaz elemeitől.




**1. ábra.** *Röntgenkép példák a négy osztályból, Normal, Covid, Lung Opacity és Viral Pneumonia. Az ezekhez tartozó Grad-CAM térképek az osztályozási folyamat szempontjából leginkább fontos régiókat emelik ki*

Az OOD-adatok kiértékelése során az első lépés a megfelelő konfidenciaszint meghatározása egy adott bemenet esetén. Az így kapott konfidenciaszint-értéket egy előre meghatározott küszöbértékkel hasonlítjuk össze, így az összehasonlítás alapján könnyen megállapítható, hogy az adott bemenet OOD- vagy ID-kategóriába tartozik-e. Ha a számított konfidenciaszint a küszöbérték alatt van, az adatot OOD-adatnak, ellenkező esetben ID-adatnak tekintjük.

A konfidenciaszintek számításához a következő módszereket használhatjuk:

– **A softmax-alapú módszerek** általában a neurális háló softmax-rétegének kimenetét veszik alapul a konfidenciaszint becsléséhez. Az ID-adatoknak magasabb a softmax-kimenete, mint az OOD-adatoknak. Így a konfidencia-pontszám definiálása a softmax-értékek függvényeként logikusnak tűnik. Az ilyen algoritmusok azonban csak egyetlen softmax-kimenettel rendelkező, klasszifikáló neurális háló esetében működnek [4–6].
– **A feature-density-alapú módszerek** a konfidenciaértéket a softmaxtól különböző, mélyebb réteg kimenetének felhasználásával számítják ki: az egyik „feature-recognition”-réteg egyikét veszik alapul. Az ötlet azon a feltételezésen alapul, hogy az OOD-adatok olyan súlyeloszlást eredményeznek mélyebb feature-recognition-rétegekben, amelyek jelentősen eltérnek az ID-minták „átlagos jellemzőeloszlásától”. A gyakorlatban tehát a sűrűségalapú algoritmusok a háló által megtanult jellemzők súlyeloszlásait veszik alapul, és valószínűségi eloszlásoknak tekintik őket. Így, a feature-eloszlások terén az alacsony sűrűségű területekre eső mintákat OOD-adatoknak tekinthetjük. A jellemzők eloszlásai, azaz az egyes jellemzők sűrűségfüggvénye egy-egy hisztogrammal közelíthető, melyeket a tanító (ID) adatok alapján kapunk. Ennek megfelelően e módszer egy histogram-based outlier score (HBOS)-eljárást valósít meg [7].

A megfelelő OOD-detektálásimódszer kiválasztása a felhasználás jellegétől függ, ideértve az adatok típusát, a neurális háló architektúráját és a modell elvárt robusztusságát [8]. Tekintettel arra, hogy a HBOS-algoritmus sokkal általánosabban alkalmazható, mint a softmax-alapú módszerek, ezt a módszert választottuk a további alkalmazási példák bemutatásához.

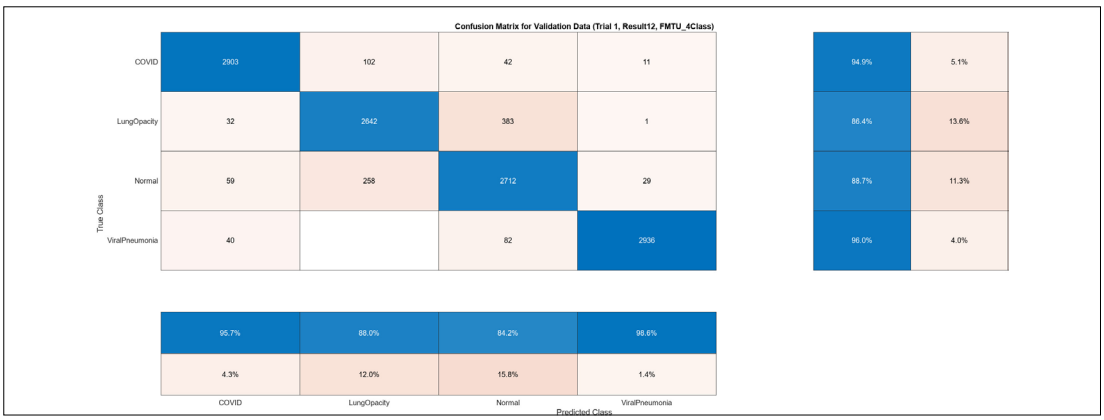
### 2.3 Megfelelő neurális háló választása

Összesen 24 221 felvételt használtunk a tanításhoz, ebből osztályonként 3058-at a validáláshoz. A Resnet18- [9] hálóarchitektúrát vettük alapul; a röntgenfelvételek, azaz a bemeneti adatok 299×299×3 méretű színeskép-formátumúak. A tanítóhalmaztól elválasztott validációs halmaz segítségével optimalizáltuk a neurálisháló-tréning paramétereit és pontosságát.

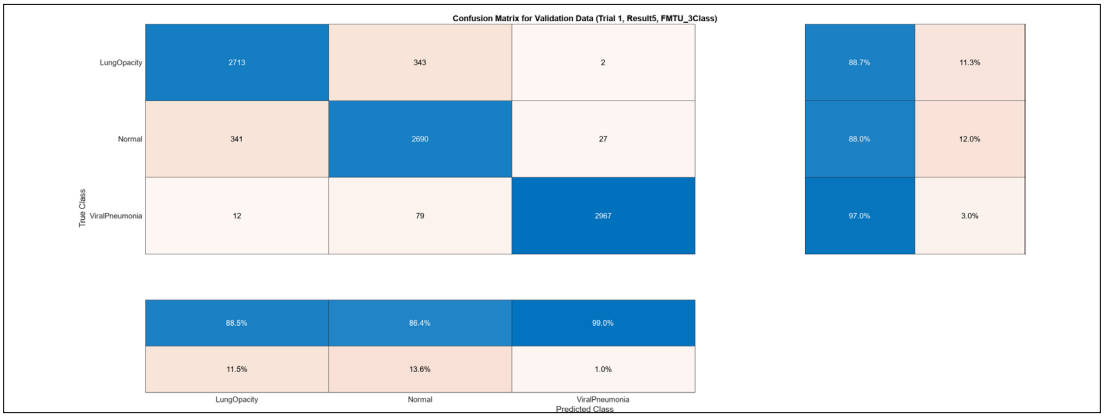
## 3. Következtetések

A Resnet18 architektúra megfelelő definiálását követően egy klasszifikáló hálót tanítottunk a teljes, 4 osztályt tartalmazó adathalmazon, valamint egy másik hálót csupán 3 osztály segítségével – kihagyva a Covid osztályt. Az 1. ábra az egyes osztályokból választott egy-egy példát mutatja, valamint a hozzájuk tartozó Grad-CAM-„hőtérképeket”, amelyek az osztályozási folyamat szempontjából leginkább fontos régiókat emelik ki.

A 2. ábrán látható konfúziós mátrix alapján világosan látszik, hogy egy jól megtanított háló mind a négy osztályt hatékonyan elkülöníti egymástól; ugyanez érvényes akkor is, ha a tanításhoz csak 3 osztályt használtunk – lásd a 3. ábrát.



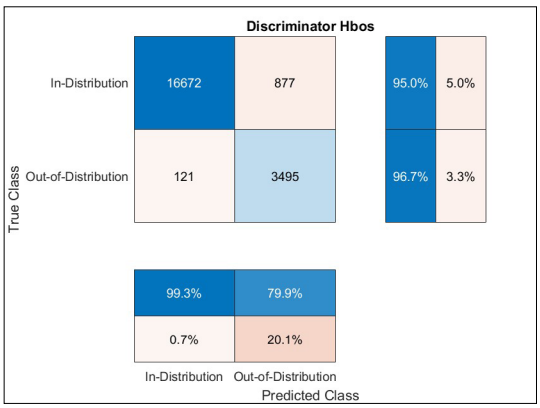

**2. ábra.** *Konfúziós mátrix a 4 osztályon tanított neurális háló esetén*




**3. ábra.** *Konfúziós mátrix a 3 osztályon tanított neurális háló esetén, mellőzve a „Covid” osztályt*

A **4. ábra** azt mutatja, hogy a HBOS-diszkriminátor jól teljesít a 3 osztályon tanított háló esetén, azaz a Covid osztály elemeit OOD-adatoknak tekinti. Igaz ugyan, hogy ez ebben a tényleges helyzetben nem meglepő, az a tény, hogy az OOD-adatok hatékonyan elkülöníthetők azoktól az adatoktól, amelyekre a hálót tanították, bővíti az AI alkalmazhatósági körét. Így a diagnózisok tekintetében a diszkriminátor jelzi, ha egy új betegséget kell feltételezni azokhoz képest, mint amelyekre az AI-eszközt ténylegesen létrehozták.

További kísérletekre lesz szükség a HBOS-diszkriminátor lehetséges finomhangolásával kapcsolatban. Továbbra is kihívást jelentő feladat ugyanis a megfelelő mély réteg automatikus kiválasztása egy neurális hálózaton belül annak érdekében, hogy a HBOS-módszer a lehető legjobb pontossággal működhessen. E cikk szerzői ennek megfelelően folytatják a kutatást, hogy optimális eljárást definiáljanak az OOD-adatok detektálására.




**4. ábra.** *HBOS-diszkriminátor konfúziós mátrixa*

## Szakirodalmi hivatkozások


[1] Yang J., Zhou K., Li Y., Liu Z.: *Generalized out-of-Distribution Detection: A Survey*. (2021,10)

[2] Chowdhury M. E. H., Rahman T., Khandakar A., Mazhar R., Kadir M. A., Mahbub Z. B., Islam K.R., Khan M. S., Iqbal A., Al-Emadi N., Reaz M. B. I., M. T. Islam: *Can AI Help in Screening Viral and COVID-19 Pneumonia?* IEEE Access, 8. (2020) 132665–132676.

[3] Rahman T., Khandakar A., Qiblawey Y., Tahir A., Kiranyaz S., Kashem S.B.A., Islam M.T., Maadeed S.A., Zughaier S.M., Khan M.S., Chowdhury M.E., *Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection Using Chest X-ray Images*. arXiv preprint arXiv:2012.02238. 2020.

[4] Shalev, Gal, Gabi Shalev, Joseph Keshet: *A Baseline for Detecting Out-of-Distribution Examples in Image Captioning*. In Proceedings of the 30th ACM International Conference on Multimedia, Lisboa Portugal. ACM, 2022. 4175–4184. https://doi.org/10.1145/3503161.3548340

[5] Shiyu Liang, Yixuan Li, R. Srikant: *Enhancing the Reliability of Out-of-Distribution Image Detection in Neural Networks*. arXiv:1706.02690 [cs.LG], August 30, 2020. http://arxiv.org/abs/1706.02690.

[6] Weitang Liu, Xiaoyun Wang, John D. Owens, Yixuan Li: *Energy-Based out-of-Distribution Detection* arXiv:2010.03759 [cs.LG], April 26, 2021. http://arxiv.org/abs/2010.03759

[7] Markus Goldstein, Andreas Dengel: *Histogram-Based Outlier Score (hbos): A Fast Unsupervised Anomaly Detection Algorithm*." KI-2012: Poster and Demo Track 9 (2012).

[8] Lee, Kimin, Kibok Lee, Honglak Lee, Jinwoo Shin: *A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks*. arXiv, October 27, 2018. http://arxiv.org/abs/1807.03888

[9] He K., Zhang X., Ren S., Sun J.: *Deep Residual Learning for Image Recognition.* (2015. 12) https://doi.org/10.48550/arXiv.1512.03385