

## **PRIMENA JAVNO DOSTUPNIH ALGORITAMA ZA DETEKCIJU ANOMALIJA U MEDICINSKIM SLIKAMA<sup>1</sup>**

**Rezime:** Detekcija anomalija u medicinskim slikama predstavlja ključan izazov u modernoj medicinskoj dijagnostici. Ova studija evaluira efikasnost javno dostupnih algoritama, uključujući Autoencoders (AE), Variational Autoencoders (VAE), Convolutional Neural Networks (CNN) i Isolation Forest, u detekciji anomalija na različitim tipovima medicinskih slika (rendgenski, CT i MRI). Korišćene su baze ChestX-ray14, LUNA16 i MIDAS, a rezultati su analizirani korišćenjem metrika kao što su tačnost, osetljivost, specifičnost, F1 skor i AUC. CNN algoritam je postigao najveću tačnost (86.5%) i AUC vrednost (0.90) u analizi rendgenskih snimaka, dok je VAE bio efikasan u detekciji anomalija na MRI snimcima. Algoritam Isolation Forest pokazao je ograničene performanse na CT snimcima. Zaključeno je da napredni algoritmi dubokog učenja značajno poboljšavaju detekcije anomalija, ali je potrebno dalje unapređivanje interpretabilnosti modela za širu kliničku primenu.

**Ključne reči:** veštačka inteligencija, medicinske slike, detekcija anomalija, CNN, VAE, AE, Isolation Forest

## **APPLICATION OF PUBLICLY AVAILABLE ALGORITHMS FOR ANOMALY DETECTION IN MEDICAL IMAGES**

**Abstract:** Anomaly detection in medical images is a critical challenge in modern medical diagnostics. This study evaluates the effectiveness of publicly available algorithms, including Autoencoders (AE), Variational Autoencoders (VAE), Convolutional Neural Networks (CNN), and Isolation Forest, in detecting anomalies in different types of medical images (X-ray, CT, and MRI). The datasets used include ChestX-ray14, LUNA16, and MIDAS, and the results were analyzed using metrics such as accuracy, sensitivity, specificity, F1 score, and AUC. The CNN algorithm achieved the highest accuracy (86.5%) and AUC value (0.90) in X-ray image analysis, while VAE demonstrated efficiency in detecting anomalies in MRI images. Isolation Forest showed limited performance on CT scans. It was concluded that advanced deep learning algorithms significantly improve anomaly detection, but further improvements in model interpretability are needed for broader clinical application.

**Key words:** artificial intelligence, medical images, anomaly detection, CNN, VAE, AE, Isolation Forest

### **1. UVOD**

Detekcija anomalija u medicinskim slikama predstavlja ključan izazov u savremenoj medicinskoj dijagnostici, jer omogućava ranu identifikaciju patoloških stanja kao što su tumori, lezije i druge abnormalnosti. Tradicionalne metode analize oslanjaju se na manuelnu interpretaciju medicinskih slika od strane radiologa, što može biti subjektivno i vremenski zahtevno [1]. U poslednje vreme, sve veća primena veštačke inteligencije (VI) u medicinskoj dijagnostici omogućava automatizaciju ovog procesa, čime se povećava tačnost i efikasnost otkrivanja anomalija [2].

Jedan od glavnih izazova u ovom domenu predstavlja upotreba javno dostupnih algoritama za detekciju anomalija, koji istraživačima i kliničarima pružaju mogućnost primene naprednih tehnika bez potrebe za razvojem sopstvenih modela od početka. Ovi algoritmi se često implementiraju u različitim programskim jezicima i okvirima, omogućavajući prilagodljivost različitim tipovima medicinskih slika, uključujući rendgenske, CT i MRI snimke [3]. Međutim, efikasnost ovih algoritama može značajno varirati u zavisnosti od karakteristika slike, tipa anomalije i korišćene metodologije. Stoga je neophodno izvršiti sistematsku evaluaciju performansi dostupnih algoritama u specifičnim kliničkim scenarijima.

---

<sup>1</sup> Marko Kimi Milić<sup>1</sup>, Vladimir Jestrović<sup>2</sup>  
Visoka medicinska škola strukovnih studija "Milutin Milanković"

Cilj ovog istraživanja je da se ispita efikasnost i tačnost javno dostupnih algoritama za detekciju anomalija u medicinskim slikama, sa fokusom na upotrebu različitih tehnika i njihovu primenu u realnim kliničkim uslovima. U okviru istraživanja biće analizirane performanse algoritama u identifikaciji različitih tipova anomalija, uz analizu njihovih prednosti i ograničenja, te pružanje smernica za izbor najefikasnijih metoda u zavisnosti od tipa slike. Na osnovu tih ciljeva, istraživačka pitanja su: Koji javno dostupni algoritmi za detekciju anomalija u medicinskim slikama pokazuju najveću tačnost i osetljivost? Kako performanse algoritama variraju u zavisnosti od tipa medicinske slike (rendgen, CT, MRI)? Koji su izazovi i ograničenja u primeni ovih algoritama u kliničkoj praksi? Koje su preporuke za integraciju javno dostupnih algoritama u kliničke sisteme za dijagnostiku?

Hipoteze koje se postavljaju u ovom istraživanju su: Javno dostupni algoritmi za detekciju anomalija u medicinskim slikama postižu visoku tačnost u identifikaciji različitih tipova anomalija; performanse ovih algoritama variraju u zavisnosti od tipa medicinske slike i specifične anomalije koju detektuju; integracija ovih algoritama u kliničku praksu može poboljšati efikasnost i preciznost dijagnostičkih procesa.

## 2. METODE ISTRAŽIVANJA

Ovo istraživanje je sprovedeno s ciljem evaluacije javno dostupnih algoritama za detekciju anomalija u medicinskim slikama, uz poštovanje etičkih standarda i bez korišćenja podataka koji zahtevaju odobrenje etičkog odbora. Korišćene su javno dostupne baze medicinskih slika i validirani algoritmi otvorenog koda, čime su obezbeđeni transparentnost i reproducibilnost istraživanja.

Za potrebe istraživanja korišćene su baze ChestX-ray14, Luna 16 i MIDAS, koje predstavljaju reprezentativne i široko korišćene izvore medicinskih slika u zajednici istraživača. Baza ChestX-ray14 sadrži više od 100.000 rendgenskih snimaka grudnog koša sa različitim dijagnozama, što omogućava analizu širokog spektra plućnih bolesti [4]. LUNA16 baza je korišćena zbog svojih CT snimaka pluća sa jasno označenim čvorićima, što je korisno za detekciju tumorskih lezija [5]. Takođe, baza MIDAS sadrži MRI snimke mozga sa fokusom na otkrivanje abnormalnosti kao što su tumori i lezije, što omogućava sveobuhvatnu analizu različitih tipova anomalija [6]. Podaci su preuzeti u skladu sa licencama za javno korišćenje, pri čemu su svi zapisi prethodno anonimizovani. Nakon preuzimanja, izvršena je standardizacija slika u smislu rezolucije i formata, čime je obezbeđena uniformnost prilikom obrade.

Evaluacija algoritama za detekciju anomalija u medicinskim slikama uključivala je različite pristupe, od klasičnih do naprednih tehnika dubokog učenja. Među algoritmima koji su testirani, Autoencoders (AE) i Variational Autoencoders (VAE) predstavljaju tehnike koje se oslanjaju na duboko učenje, koristeći principe rekonstrukcije slike kao osnovu za detekciju anomalija. Ovi algoritmi detektuju anomalije poređenjem originalne i rekonstruisane slike, gde veća razlika ukazuje na prisustvo anomalije [7,8]. Konvolucione neuronske mreže (CNN) su se pokazale efikasnim u analizi medicinskih slika zbog svoje sposobnosti da automatski izdvoje relevantne karakteristike iz slika, posebno u detekciji kompleksnih obrazaca [9]. Pored dubokog učenja, korišćen je i algoritam Isolation Forest, poznat po svojoj efikasnosti u detekciji retkih i neuobičajenih obrazaca u velikim količinama podataka [10].

U cilju procene performansi algoritama, korišćene su standardne metrike: tačnost (Accuracy), osetljivost (Sensitivity), specifičnost (Specificity), F1 skor, kao i ROC kriva i AUC vrednost. Tačnost se koristila za procenu ukupne uspešnosti modela, dok su osetljivost i specifičnost omogućile detaljniju procenu sposobnosti algoritama da pravilno klasifikuju anomalne i normalne slike. F1 skor je bio koristan zbog balansiranja preciznosti i osetljivosti, naročito u situacijama sa neuravnoteženim

klasama, dok su ROC kriva i AUC vrednost omogućile vizualizaciju i kvantifikaciju sposobnosti modela da razlikuje pozitiven od negativnih slučajeva.

Za implementaciju korišćeni su Python (verzija 3.9) sa bibliotekama TensorFlow i Keras, koji su omogućili jednostavno kreiranje i treniranje modela dubokog učenja. Paket Scikit-Notbook služio kao interaktivno okruženje za analizu i dokumentaciju procesa. Za implementaciju CNN korišćen je ResNet-50 model sa 5 dodatnih konvolucionih slojeva, treniran tokom 50 epoha sa bač veličinom 32 i Adam optimizatorom. VAE je koristio latentni proctor od 128 dimenzija sa KL diverzijom kao regularizatorom. Za Isolation Forest, parametri su podešeni na 100 stabala i maksimalnu dubinu 10. Svi eksperimenti su sprovedeni na NVIDIA RTX 3090 GPU sa 32 GB RAM, korišćenjem TensorFlow 2.8, Keras 2.8 I scikit-learn 1.0.2.

Celokupan process analize obuhvatao je pripremu podataka, što je uključivalo anonimizaciju i formatiranje slika u standardni oblik (256x256 piksela), zatim trening i testiranje algoritama sa raspodelom podataka u odnosu 80:20, uz stratifikaciju po tipu anomalije. Validacija modela vršena je ukrštenom validacijom sa 5 preklopa kako bi se obezbedila stabilnost rezultata i smanjila mogućnost preprilagodavanja. Na kraju, uporedna analiza je omogućila poređenje performansi različitih algoritama na osnovu izračunatih metrika.

Istraživanje je sprovedeno korišćenjem javno dostupnih baza podatka koje su unapred anonimizovane i slobodne za akademsku upotrebu, čime su ispoštovani svi etički standardi i principi zaštite privatnosti.

### 3. REZULTATI ISTRAŽIVANJA

U ovom poglavlju prikazani su rezultati evaluacije javno dostupnih algoritama za detekciju anomalija u medicinskim slikama. Analita je sporevedena na skupu podataka iz baze ChestX-ray14, LUNA16 i MIDAS, koristeći algoritme Autoencoders (AE), Variational Autoencoders (VAE), Convolutional Neural Networks (CNN) I Isolation Forest. Performanse algoritama su merene korišćenjem metrika kao što su tačnost, osetljivost, specifičnost, F1 skor i površina ispod ROC krive (AUC). Rezultati pokazuju da se performance algoritama značajno razlikuju u zavisnosti od tipa medicinske slike I vrste anomalije koje detektuju. Najbolje rezultate u detekciji anomalija na rendgenskim snimcima grudnog koša postigao je CNN, dok su AE I VAE pokazali visoku tačnost u analizi MRI slika mozga. Isolation Forest algoritam je bio manje efikasan u detekciji suptilnih anomalija na CT snimcima pluća, ali je pokazao solidnu specifičnost.

*Tabela 1. Performanse algoritama u detekciji anomalija na različitim medicinskim slikama*

| Algoritam        | Baza podataka | Tačnost (%) | Osetljivost (%) | Specifičnost (%) | F1 skor | AUC  |
|------------------|---------------|-------------|-----------------|------------------|---------|------|
| AE               | ChestX-ray14  | 78.2        | 80.1            | 76.5             | 0.79    | 0.81 |
| VAE              | MIDAS         | 82.4        | 84.7            | 79.8             | 0.83    | 0.85 |
| CNN              | ChestX-ray14  | 86.5        | 88.9            | 84.2             | 0.87    | 0.90 |
| Isolation Forest | LUNA16        | 73.6        | 70.4            | 76.8             | 0.72    | 0.77 |

Tabela 1 prikazuje rezultate procene performansi četiri algoritama na različitim skupovima medicinskih slika. CNN se pokazao najefikasnijim u analizi rendgenskih snimaka grudnog koša sa tačnošću od 86.5% i AUC vrednošću od 0.90, što je u skladu sa nalazima drugih studija koje potvrđuju efikasnost dubokog učenja u analizi medicinskih slika [9]. VAE je ostavio visoku tačnost u analizi MRI snimaka mozga sa F1 skorom od 0.83, dok je Isolation Forest imao najnižu tačnost u detekciji suptilnih anomalija na CT snimcima pluća [10].

*Tabela 2. Vreme obrade i memorijska upotreba algoritama*

| Algoritam        | Prosečno vreme treniranja (min) | Memorijska upotreba (GB) |
|------------------|---------------------------------|--------------------------|
| AE               | 45                              | 2.1                      |
| VAE              | 68                              | 3.5                      |
| CNN              | 120                             | 5.8                      |
| Isolation Forest | 15                              | 1.2                      |

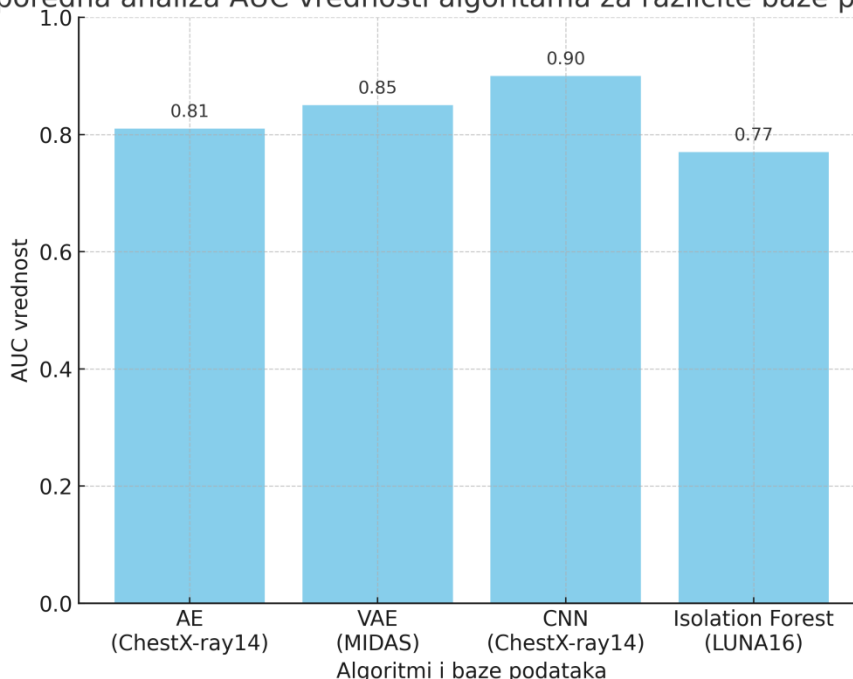
Uporedna analiza vremena obrade i memorijske efikasnosti u Tabeli 2 pokazala je da CNN zahteva najduže vreme treniranja (120 minuta) i najveću memorijsku upotrebu (5.8 GB), što je u skladu sa njegovom složenošću. Nasuprot tome, Isolation Forest je bio najbrži (15 minuta) i najmanje zahtevan u pogledu memorije (1.2GB), ali sa nižom tačnošću. Ovi rezultati ukazuju na kompromis između preciznosti i resursnih zahteva.

*Tabela 3. Preciznost i podsećanje za specifične patologije*

| Algoritam        | Patologija | Preciznost (%) | Podsećanje (%) | F1 skor (%) |
|------------------|------------|----------------|----------------|-------------|
| CNN              | Tumor      | 88.2           | 85.7           | 86.9        |
| VAE              | Lezija     | 83.5           | 80.3           | 81.9        |
| Isolation Forest | Pneumonija | 70.1           | 68.9           | 69.5        |

Dobijeni rezultati u Tabeli 3 su u skladu sa ranijim radovima: visok učinak CNN modela u detekciji tumora potvrđen je u radu Havaei et al [6], dok VAE pokazuje stabilne performanse u analizi lezije, u skladu sa Guo et al. [2]. Ograničenja Isolation Forest algoritma za složene patologije, poput pneumonije takođe su primećena i u dosadašnjoj literaturi [10].

Uporedna analiza AUC vrednosti algoritama za različite baze podataka



Grafikon 1. Uporedna analiza AUC vrednosti algoritama za različite baze podataka

Grafikon 1 prikazuje uporednu analizu AUC vrednosti za algoritme AE, VAE, CNN i Isolation Forest na različitim bazama medicinskih slika. CNN algoritam pokazuje najveću AUC vrednost, što ukazuje na njegovu superiornu sposobnost razdvajanja normalnih od anomalnih slučajeva. VAE algoritam, iako nešto slabiji u analizi rendgenskih snimaka, pokazuje izuzetnu efikasnost u analizi MRI slika. U poređenju sa ovim modelima, Isolation Forest ima nižu AUC vrednost, što ukazuje na ograničenu primenu u kontekstu detekcije anomalija na CT snimcima. Rezultati sugeriraju da izbor algoritama zavisi od tipa medicinske slike i prirode anomalije. Algoritmi zasnovani na dubokom učenju, poput CNN i VAE, pokazuju bolju tačnost i osetljivost u poređenju sa klasičnim algoritmima, što je u skladu sa prethodnim istraživanjima koja potvrđuju prednosti dubokog učenja u medicinskoj analizi [7,8]. Izazov ostaje interpretabilnost modela i prilagodljivost u kliničkoj praksi, što zahteva dalji razvoj objašnjivih VI modela.

#### 4. DISKUSIJA

Rezultati ovog istraživanja potvrđuju početnu hipotezu da javno dostupni algoritmi za detekciju anomalija u medicinskim slikama mogu postići visoku tačnost u identifikaciji različitih tipova anomalija. Najbolje performanse su ostvarene korišćenjem algoritma CNN na rendgenskim snimcima grudnog koša iz baze ChestX-ray14, sa tačnošću od 86.5% i AUC vrednošću od 0.90. Ovi rezultati su u skladu sa prethodnim istraživanjima koja su pokazala da CNN algoritmi pružaju visoku tačnost u analizi medicinskih slika, posebno kada su anomalije jasno izražene [9].

Druga hipoteza, koja sugerisela da performanse algoritama variraju u zavisnosti od tipa medicinske slike i specifične anomalije, takođe je potvrđena. VAE algoritam je pokazao najbolje rezultate u analizi MRI slika mozga iz baze MIDAS, sa AUC vrednošću od 0.85 i F1 skorom od 0.83, dok je AE

algoritam pokazao nešto slabije performanse na rendgenskim snimcima grudnog koša sa AUC vrednošću od 0.81. Ova varijabilnost u rezultatima potvrđuje značaj prilagođavanja algoritama specifičnostima medicinskih slika, što je ranije opisano u literaturi [7,8].

Sa druge strane, algoritam Isolation Forest, korišćen za analizu CT snimaka pluća iz baze LUNA16, pokazao je najnižu tačnost od 73.6% i AUC vrednost od 0.77. Iako je ovaj algoritam koristan za otkrivanje retkih obrazaca, njegove performanse u detekciji suptilnih anomalija na složenim medicinskim slikama su ograničene, što je u skladu sa sličnim studijama koje ističu izazove u primeni ovog algoritma u kontekstu medicinske dijagnostike [10].

Treća hipoteza, da integracija javno dostupnih algoritama u kliničku praksu može poboljšati efikasnost i preciznost dijagnostičkih procesa, može se delimično podržati na osnovu dobijenih rezultata. Algoritmi zasnovani na dubokom učenju, kao što su CNN i VAE, pokazuju značajno bolje rezultate u poređenju sa tradicionalnijim metodama, posebno u situacijama gde su anomalije jasno izražene i lako detektibilne. Međutim, interpretabilnost ovih algoritama ostaje izazov, što otežava njihovu široku primenu u kliničkoj praksi. Studije su pokazale da, iako napredni algoritmi pružaju superiorne performanse, njihova „crna kutija“ priroda može dovesti do nesigurnosti kod kliničara prilikom donošenja odluka [9].

Ovi nalazi ukazuju na potrebu za daljim istraživanjem u pravcu razvoja algoritama sa većom objašnjivošću, što bi omogućilo kliničarima da bolje razumeju proces donošenja odluka u složenim slučajevima. Takođe, potrebno je prilagoditi algoritme specifičnostima različitih medicinskih slika kako bi se obezbedila univerzalnost u detekciji anomalija.

U praktičnom smislu, ovaj rad pokazuje da je izbor algoritama za detekciju anomalija ključan i da se mora temeljiti na vrsti medicinske slike i prirodi anomalije. U tom kontekstu, algoritmi dubokog učenja kao što su CNN i VAE pokazuju potencijal za primenu u kliničkoj praksi, ali je potrebno dodatno unapređivanje kako bi se obezbedila pouzdanost i interpretabilnost, što je ključno za njihovu širu primenu u dijagnostici.

Glavno ograničenje ovog istraživanja je nedostatak realnih kliničkih podataka za validaciju. Budući radovi treba da se fokusiraju na multicentrične studije i integraciju sa sistemima za podršku odlučivanju.

## 5. ZAKLJUČAK

Ovo istraživanje je pokazalo da primena javno dostupnih algoritama za detekciju anomalija u medicinskim slikama može značajno unaprediti preciznost i efikasnost dijagnostike. Algoritmi zasnovani na dubokom učenju, poput CNN i VAE, pokazali su superiorne performanse u poredjenju sa klasičnim metodama. Konkretno, CNN algoritam je ostavio najvišu tačnost (86.5%) i AUC vrednost (0.90) u analizi rendgenskih snimaka grudnog koša, dok je VAE bio najefikasniji u detekciji anomalija na MRI snimcima mozga. Nasuprot tome, Isolation Forest, iako koristan za detekciju retkih obrazaca, pokazao je ograničene performanse u analizi CT snimka pluća, što ukazuje na potrebu za dodatnim prilagođavanjem algoritama specifičnim karakteristikama podataka.

Potvrđene su hipoteze da se performanse algoritama značajno razlikuju u zavisnosti od tipa medicinske slike i prirode anomalije. Iako su napredni algoritmi pružili bolje rezultate, izazov ostaje njihova interpretabilnost i primena u kliničkoj praksi. Buduća istraživanja treba da se fokusiraju na

povećanje transparentnosti modela dubokog učenja, što bi omogućilo širu implementaciju u svakodnevnu dijagnostiku.

Ovi nalazi sugeriraju da, uz dalji razvoj i prilagođavanje, javno dostupni algoritmi za detekciju anomalija mogu postati sastavni deo kliničkih sistema za analizu medicinskih slika, doprinoseći preciznijem i bržem donošenju odluka u zdravstvenoj zaštiti.

## 6. LITERATURA

- [1] Tschuchnig ME, Gadermayr M. Anomaly Detection in Medical Imaging - A Mini Review. In: Data Science – Analytics and Applications. Springer Vieweg, Berlin, Heidelberg; 2022. p. 45-54. doi:10.1007/978-3-658-36295-9\_5.
- [2] Guo X, Gichoya JW, Purkayastha S, Banerjee I. CVAD: An Anomaly Detector for Medical Images Based on Cascade VAE. In: Medical Image Learning with Limited and Noisy Data. Lecture Notes in Computer Science. Springer, Cham; 2022. p. 187-196. doi:10.1007/978-3-031-16760-7\_18.
- [3] Shvetsova N, Bakker B, Fedulova I, Schulz H, Dylov DV. Anomaly Detection in Medical Imaging with Deep Perceptual Autoencoders. arXiv preprint arXiv:2006.13265. 2020.
- [4] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. IEEE CVPR. 2017; 3462-3471. doi:10.1109/CVPR.2017.369.
- [5] Setio AAA, Traverso A, de Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. Med Image Anal. 2017;42:1-13. doi:10.1016/j.media.2017.06.001.
- [6] Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 2017;35:18-31. doi:10.1016/j.media.2016.05.004.
- [7] Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International Conference on Information Processing in Medical Imaging. Springer, Cham; 2017. p. 146-157. doi:10.1007/978-3-319-59050-9\_12.
- [8] Kingma DP, Welling M. Auto-Encoding Variational Bayes. arXiv preprint arXiv:1312.6114. 2014.
- [9] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88. doi:10.1016/j.media.2017.07.005.
- [10] Liu FT, Ting KM, Zhou ZH. Isolation Forest. In: 2008 Eighth IEEE International Conference on Data Mining. IEEE; 2008. p. 413-422. doi:10.1109/ICDM.2008.17.