

Učinkovitost umjetne inteligencije u razlikovanju malignih od benignih ultrazvučnih osobina tumora dojki

Filip Marković¹, Damir Štimac^{1,2}

¹ Medicinski fakultet Osijek, Osijek

² Nacionalna memorijalna bolnica „Dr. Juraj Njavro“ Vukovar

Corresponding author: Filip Marković, email: fmarkovic24@gmail.com

DOI: <https://doi.org/10.55378/rv.49.1.2>

Sažetak

U ženskoj populaciji karcinom dojki je najčešća maligna bolest i mortalitet mu je 20 % zbog čega predstavlja veliki društveni i zdravstveni problem, osobito u razvijenim zemljama. Rano otkrivanje karcinoma dojke igra ključnu ulogu u poboljšanju ishoda liječenja. Ultrazvučna dijagnostika predstavlja važan alat za prepoznavanje i razlikovanje sumnjivih lezija, no precizna evaluacija može biti izazovna zbog subjektivnosti interpretacije. U tom kontekstu, umjetna inteligencija (UI), uključujući duboko učenje nude mogućnost standardizacije i unaprijeđenja dijagnostičke točnosti. Ova retrospektivna studija provedena je na Odjelu za radiologiju Nacionalne memorijalne bolnice „Dr. Juraj Njavro“ u Vukovaru, koristeći 257 ultrazvučnih slika pacijentica koje su u razdoblju od studenog 2021. do veljače 2024. obavile ultrazvučni pregled i patohistološku analizu iglenom biopsijom. Prikupljene slike prilagođene su za analizu pomoću neuronske mreže *DenseNet*. Cilj ovog istraživanja bio je ispitati učinkovitost neuronske mreže *DenseNet* u analizi ultrazvučnih slika tumora dojki, s naglaskom na razlikovanje malignih od benignih tvorbi i mogućnost smanjenja broja nepotrebnih invazivnih dijagnostičkih postupaka. Dijagnostička učinkovitost modela procijenjena je usporedbom njegovih rezultata s patohistološkim nalazima, a korišteni su standardni statistički testovi. Model je pokazao točnost od 73%, AUC (area under the curve) vrijednost od 0.78, specifičnost od 68.8%, osjetljivost od 80% i F1-score od 70%.

Ključne riječi: tumor dojke; ultrazvuk dojke; umjetna inteligencija

Uvod

Karcinom dojke najčešća je maligna bolest kod žena, s visokim morbiditetom i mortalitetom, posebno u razvijenim zemljama (1). Pravovremeno i točno razlikovanje benignih od malignih lezija dojke ključno je za učinkovito liječenje i smanjenje broja nepotrebnih invazivnih zahvata. Ultrazvuk je često korištena metoda u dijagnostici promjena u dojci, pružajući detaljan uvid u morfologiju i strukturu lezija. No, interpretacija ultrazvučnih slika može biti subjektivna i ovisi o iskustvu radiologa, što može dovesti do varijabilnosti u dijagnozi.

Razvoj umjetne inteligencije (UI) u medicini, posebno dubokog učenja i neuronskih mreža, pruža nove mogućnosti za automatizaciju i unaprijeđenje točnosti dijagnostike u radiologiji. Neuronska mreža *DenseNet*, poznata po svojoj gusto povezanoj arhitekturi, omogućuje prijenos informacija kroz sve slojeve mreže i precizno prepoznavanje karakteristika tumora. Ova arhitektura omogućuje pouzdano prepoznavanje malignih i benignih karakteristika na temelju medicinskih slika, čime se olakšava donošenje kliničkih odluka i smanjuje potreba za dodatnim dijagnostičkim postupcima.

U ovoj studiji istražena je primjena *DenseNet* mreže u analizi ultrazvučnih slika tumora dojke, s ciljem procjene njene točnosti i potencijala za kliničku primjenu u razlikovanju malignih od benignih lezija, što bi moglo pridonijeti smanjenju broja nepotrebnih invazivnih postupaka te poboljšanju dijagnostičkog procesa.

Materijali i metode

Ovaj rad proveden je kao retrospektivna studija temeljem sekundarnih podataka, s ciljem analize mogućnosti primjene dubokog učenja u razlikovanju benignih i malignih tumora dojki. Studija je provedena na Odjelu za radiologiju Nacionalne memorijalne bolnice „Dr. Juraj Njavro“ u Vukovaru, a uključuje 257 ultrazvučnih slika dojki pacijentica koje su obavile ultrazvučni pregled i patohistološku dijagnostiku iglenom biopsijom između studenog 2021. i veljače 2024. Ovi podaci prikupljeni su iz bolničkog informacijskog sustava (BIS) te su ultrazvučne slike prilagođene za analizu u modelu umjetne inteligencije (UI), s obradom koja uključuje označavanje područja tumora i prilagođavanje veličine piksela kako bi slike bile kompatibilne za obradu



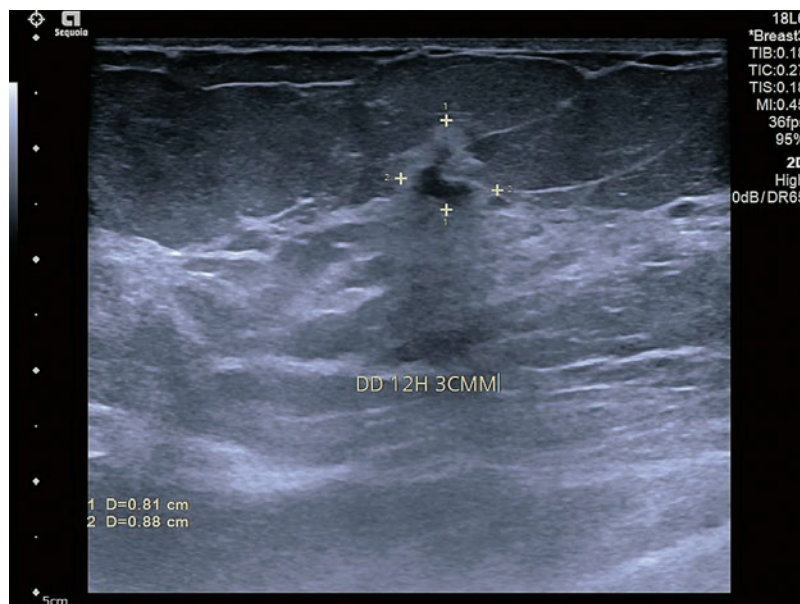
Slika 1. UZV benignog tumora dojke ispravno klasificiran umjetnom inteligencijom

(Slike 1. i 2.). Pacijentice su potpisale informirani pristanak obrade podataka i rad je odobrilo etičko povjerenstvo NMB „Dr. Juraj Njavro“ Vukovar.

Za analizu je korišten UI model temeljen na neuronskoj mreži *DenseNet-121*. Ovaj model, s 121 slojem, prethodno je treniran na *ImageNet* skupu podataka, pri čemu su zadnja četiri sloja modela otključana za fino podešavanje. Podaci su podijeljeni na tri skupine: 80% slika korišteno je za trening, 10% za validaciju, a preostalih 10% za testiranje modela. „Trening“ skup koristi se za inicijalno učenje modela, gdje model uči prepoznavati obrasce u slikama. „Validacijski“ skup pomaže u procjeni općenitosti modela i sprječava prekomjerno učenje, dok je „testni“ skup rezerviran za konačnu provjeru modela na potpuno novim podacima, što simulira rad modela u stvarnom kliničkom scenariju. Tijekom treniranja, slike

su prolazile kroz model u epohama (jedan prolaz kroz cjelokupni skup za treniranje), a model je treniran tijekom 15 epoha uz primjenu metode ranog zaustavljanja (engl. *early stopping*) kako bi se optimizirao proces učenja i smanjio rizik od prenaučivosti. Praćene su vrijednosti gubitka (engl. *validation loss*) i točnosti na validacijskom skupu (engl. *validation accuracy*) kako bi se prilagodio proces treniranja, dok je brzina učenja postavljena na 0,01, broj koji određuje koliko brzo model ažurira težine tijekom treniranja.

Za evaluaciju modela korišteni su standardni dijagnostički testovi, uključujući točnost (engl. *accuracy*), osjetljivost (engl. *sensitivity*) = $TP/(TP + FN)$, preciznost (engl. *precision*) = $TP/(TP + FP)$, specifičnost = $TN/(TN + FP)$ (specificity), *F1-score* = $2 \times (Precision \times Sensitivity) / (Precision + Sensitivity)$ te vrijednost površine ispod



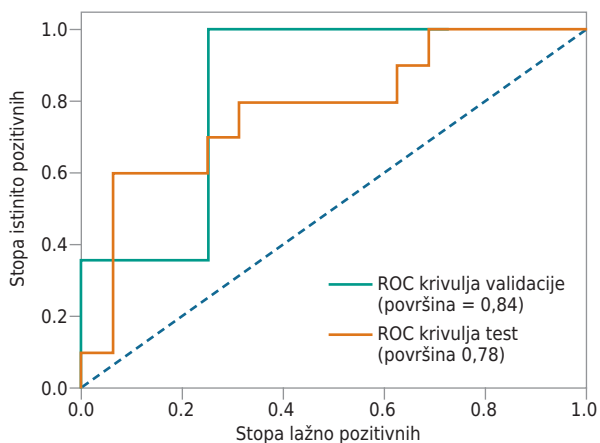
Slika 2. UZV malignog tumor dojke ispravno klasificiran umjetnom inteligencijom

krivulje (AUC). Osjetljivost je određena omjerom istinski pozitivnih (TP) rezultata i zbroja istinski pozitivnih i lažno negativnih (FN), dok je preciznost omjer istinski pozitivnih i zbroja istinski pozitivnih i lažno pozitivnih rezultata (FP). Specifičnost je definirana kao omjer istinski negativnih (TN) i zbroja istinski negativnih i lažno pozitivnih (FP) rezultata. F1-mjera izračunata je kao harmonijska sredina između preciznosti i osjetljivosti kako bi se osigurao uravnotežen pokazatelj uspješnosti modela.

Rezultati

Testiranje modela radilo se na 26 slika, od kojih je 16 slika prikazivalo benigne a 10 maligne lezije. AUC testnog skupa bio je 0.78 (Slika 3.).

Osjetljivost modela za maligne lezije bila je 80%, a točnost 73%. Nešto lošiji rezultati bili su rezultati preciznosti (62%) i F1-score (70%). Izračunata specifičnost iz ovih podataka iznosila je 68.8% za skupinu podataka malignih lezija.



Slika 3. Krivulja operativnih karakteristika prijemnika (ROC)

Tablica 1. Tablica klasifikacija

	Preciznost	Osjetljivost	F1- score	Broj slika
Benigno	0.85	0.69	0.76	16
Maligno	0.62	0.80	0.70	10
Točnost			0.73	26

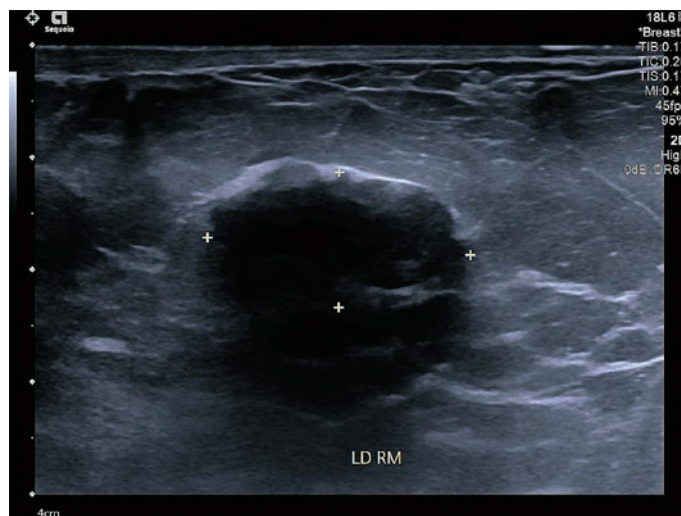
U klasifikaciji tumora, model je ispravno identificirao 11 benignih i 8 malignih uzoraka (Tablica 2., Slika 1., Slika 2., Slika 4., Slika 5.).

Tablica 2. Matrica zabune

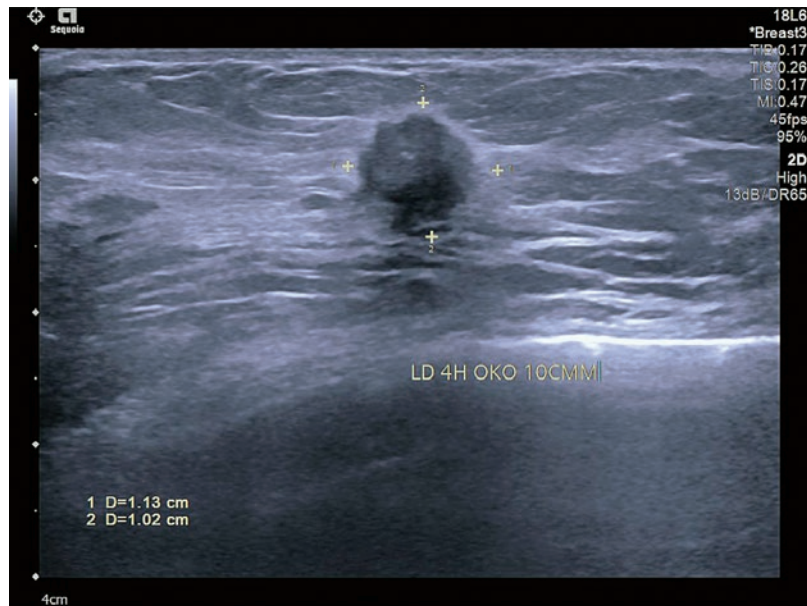
	Benigno	Maligno
Stvarno benigno	11	5
Stvarno maligno	2	8

Rasprava

U našem istraživanju obuhvaćena su 126 benigna i 131 maligni tumor, što je dobar preduvjet za kvalitetno treniranje modela zbog balansiranja ispitivanih grupa. U ostalim istraživanjima raspored grupa znao je varirati pa u jednom istraživanju raspored je bio ujednačen s 93 maligne i 102 benigne lezije (2) dok neka druga istraživanja imaju velike razlike između grupa kao npr. 575 benignih i 326 malignih promjena (3). To znači da se program može istrenirati i na nebalansiranim grupama, ali potrebno je obratiti pažnju na ostale hiperparametre treniranja i evaluacije. Jedna je od čestih zabluda potreba za velikim brojem slika kako bi se pravilno i što bolje istrenirao UI model. Veći broj slika je prednost u treningu i provjeri učinkovitosti, ali iz našeg istraživanja vidimo da se i sa 200 slika može napraviti model i istrenirati da pokazuje vjerodostojne rezultate. Kvaliteta slika korištenih za treniranje UI modela igra ključnu ulogu u njegovoj preciznosti i sposobnosti razlikovanja različitih karakteristika. Model treniran na slikama visoke kvalitete, koje su jasno definirane i pravilno označene, ima veću šansu za postizanje boljih rezultata. Niska rezolucija, šum ili loša kontrastnost mogu otežati modelu prepoznavanje relevantnih obilježja, što izravno utječe na njegovu učinkovitost i pouzdanost u stvarnim scenarijima. U ovom istraživanju korištene su ultrazvučne slike u B-modu dostupne na odjelu, različitih rezolucija i kontrasta (Slika 6.). U istraživanju Starena i suradnika, sve pacijentice se snimalo u istom položaju i isključivalo nejasne slike UZV dojke, kako bi se osigurao dobar trening i učinkovitost UI modela (4). Istraživanja su isto tako koristila različite pristupe od isključivo B-mod ultrazvučnih slika dojke do dodatnih mogućnosti npr. power Doppler analize ili ultrazvuka s kontrastom za poboljšanje klasifikacije malignih promjena dojke (5, 4). Testiranjem modela dobili su se dobri, ali ne i izvrsni rezultati, što se vidi iz AUC testnog skupa koji je iznosio je 0.78. Model ima dobru sposobnost razlikovanja između benignih i malignih tumora. Dobivena AUC vrijednost testnog skupa je slična drugim istraživanjima, gdje je kod korištenja multimodalnih ultrazvuka AUC bio 0.85 ili gdje je kod korištenja 3D power Doppler ultrazvuka najmanja vrijednost bila 0.782.



Slika 4. Benigni tumor klasificiran kao maligni.

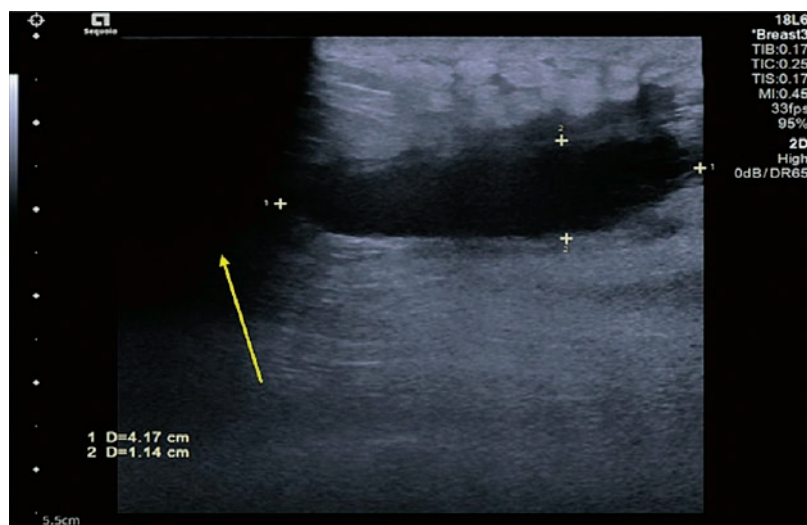


Slika 5. Maligni tumor klasificiran kao benigni

U jednom istraživanju autori su koristili multimodalni UZV u poboljšanju točnosti procjene sentinelnih limfnih čvorova u slučajevima raka dojke (7). U tom istraživanju za dvodimenzionalni (2D) UZV, AUC je iznosio 0.678 što je znatno manje od dobivene AUC u ovome istraživanju, ali bitno je naglasiti da UI modeli nisu jednako klasificirali (6, 7).

Točnost ovog modela na testnom skupu bila je 73% što je vjerojatno zbog različitih kvaliteta i rezolucija slika. Točnost u istraživanjima znatno varira od točnosti kao što je 70.5% na B-mod ultrazvuku pa sve do točnosti od preko 90% korištenjem multimodalnih ultrazvuka (6, 8, 9). U nekim istraživanjima koja su koristila modulacije slike prije prolaska kroz model točnost je iznosila čak i 96% (10). Iako još uvijek postoje modeli istovjetne točnosti kao što je naš model, ipak mislimo da ima prostora za poboljšanje. Primjenom dodatnih ultrazvučnih tehnika ili dodatnim procesiranjem slika očekujemo da bi točnost ovog modela rasla. Preciznost u ovom modelu bila je 62%.

U istraživanju preciznost na analizi 3D ultrazvučne slike iznosila je 84.7% dok na običnoj ultrazvučnoj slici 75.9%, što govori da UI modeli još uvijek nisu dovoljno precizni u određivanju malignih lezija i to u tom su segmentu potrebna daljnja istraživanja (8). Ovaj model pokazuje znatno nižu specifičnost (68.8%) od modela koji su analizirali slike multimodalnih UZV-a, gdje je specifičnost bila 92.4%, nešto manju specifičnost od modela koji su koristili analizu slika 3D UZV (75.6%), ali ima veću specifičnost od modela koji je analizirao 2D UZV slike (64.9%) (6, 8, 9). U usporedbi s modelom koji je analizirao iste slike ovaj model točnije prepoznaje benigne slučajeve. Ovaj je model osjetljiviji (80%) u komparaciji s modelom koji je koristio multimodalne UZV slike (77.4%) (6). Model testiran na 3D UZV slikama pokazuje najveću osjetljivost (90.4%), a jedan od modela koji je istraživao samo B-mod UZV slike ima nešto višu osjetljivost od našeg (81.7%) (8, 9). Model iz ovog istraživanja konkurentan je u usporedbi s drugim modelima, ali model na 3D UZV slikama pokazuje



Slika 6. UZV slika sa šumom iz istraživanja

znatno bolju osjetljivost što znači da postoji dosta prostor za napredovanje. *F1-score* (0.70) je niži u usporedbi s drugim istraživanjima, gdje istraživanje s nebalansiranim grupama ima *F1-score* od 0.87, a istraživanje s približnom istom razlikom u broju malignih i benignih tumora ima *F1-score* 0.97 (3, 10).

Na temelju provedenog istraživanja, može se zaključiti da je razvijeni model učinkovit u razlikovanju benignih od malignih tumora dojke. Iako postoje određena ograničenja, model pokazuje potencijal za unaprjeđenje dijagnostičkog procesa UZV-a dojke. Iako ne može potpuno eliminirati upotrebu invazivnih dijagnostičkih metoda, rezultati ukazuju na to da može značajno poboljšati razlikovanje benignog od malignog tumora. Daljnje istraživanje i optimizacija modela, uključujući primjenu dodatnih tehnika i poboljšanje kvalitete slika, mogla bi dodatno povećati njegovu točnost i pouzdanost.

References

- Ivan Damjanov, Sven Seiwert, Stanko Jukić, Marin Nola, Patologija, 5 izd. Zagreb: Medicinska naklada; 2018.
- Xu Z, Wang Y, Chen M, Zhang Q. Multi-region radiomics for artificially intelligent diagnosis of breast cancer using multimodal ultrasound. *Comput Biol Med.* 2022 Oct;149:105920. doi: 10.1016/j.combiomed.2022.105920. Epub 2022 Aug 6. PMID: 35986969.
- Wei Q, Yan YJ, Wu GG, Ye XR, Jiang F, Liu J, Wang G, Wang Y, Song J, Pan ZP, Hu JH, Jin CY, Wang X, Dietrich CF, Cui XW. The diagnostic performance of ultrasound computer-aided diagnosis system for distinguishing breast masses: a prospective multicenter study. *Eur Radiol.* 2022 Jun;32(6):4046-4055. doi: 10.1007/s00330-021-08452-1. Epub 2022 Jan 23. PMID: 35066633.
- Staren ED, O'Neill TP. Breast ultrasound. *Surg Clin North Am.* 1998 Apr;78(2):219-35. doi: 10.1016/s0039-6109(05)70310-8. PMID: 9602844.
- G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243. keywords: {Training;Convolution;Network architecture;Convolutional codes;Neural networks;Road transportation},
- Xu Z, Wang Y, Chen M, Zhang Q. Multi-region radiomics for artificially intelligent diagnosis of breast cancer using multimodal ultrasound. *Comput Biol Med.* 2022 Oct;149:105920. doi: 10.1016/j.combiomed.2022.105920. Epub 2022 Aug 6. PMID: 35986969.
- Su S, Ye J, Ke H, Zhong H, Lyu G, Xu Z. Multimodal ultrasound imaging: a method to improve the accuracy of sentinel lymph node diagnosis in breast cancer. *Front Oncol.* 2024 Mar 25;14:1366876. doi: 10.3389/fonc.2024.1366876. PMID: 38590661; PMCID: PMC10999636.
- Xiao X, Gan F, Yu H. Tomographic Ultrasound Imaging in the Diagnosis of Breast Tumors under the Guidance of Deep Learning Algorithms. *Comput Intell Neurosci.* 2022 Feb 28;2022:9227440. doi: 10.1155/2022/9227440. PMID: 35265119; PMCID: PMC8901319.
- Ma H, Tian R, Li H, Sun H, Lu G, Liu R, Wang Z. Fus2Net: a novel Convolutional Neural Network for classification of benign and malignant breast tumor in ultrasound images. *Biomed Eng Online.* 2021 Nov 18;20(1):112. doi: 10.1186/s12938-021-00950-z. PMID: 34794443; PMCID: PMC8600702.
- Chen H, Ma M, Liu G, Wang Y, Jin Z, Liu C. Breast Tumor Classification in Ultrasound Images by Fusion of Deep Convolutional Neural Network and Shallow LBP Feature. *J Digit Imaging.* 2023 Jun;36(3):932-946. doi: 10.1007/s10278-022-00711-x. Epub 2023 Jan 31. PMID: 36720840; PMCID: PMC10287618.

The Efficiency of Artificial Intelligence in Distinguishing Malignant from Benign Ultrasound Features of Breast Tumors

Abstract

In the female population, breast cancer is the most common malignant disease, with a mortality rate of 20%, making it a significant social and health problem, especially in developed countries. Early detection of breast cancer plays a crucial role in improving treatment outcomes. Ultrasound diagnostics represent a significant tool for identifying and differentiating suspicious lesions, but accurate evaluation can be challenging due to the subjectivity of interpretation. In this context, artificial intelligence (AI) and deep learning offer the potential for standardization and enhancement of diagnostic accuracy. This retrospective study was conducted at the Radiology Department of the National Memorial Hospital "Dr. Juraj Njavro" in Vukovar, using 257 ultrasound images from patients who underwent ultrasound examinations and pathohistological analysis via needle biopsy from November 2021 to February 2024. The collected images were adjusted for analysis using the DenseNet neural network. The aim of this research was to examine the effectiveness of the DenseNet neural network in analyzing ultrasound images of breast tumors, focusing on distinguishing malignant from benign formations and the potential to reduce the number of unnecessary invasive diagnostic procedures. The diagnostic efficacy of the model was assessed by comparing its results with pathohistological findings, utilizing standard statistical tests. The model demonstrated an accuracy of 73%, an AUC (area under the curve) value of 0.78, a specificity of 68.8%, a sensitivity of 80%, and an *F1-score* of 70%.

Keywords: artificial Intelligence; breast tumor; breast ultrasound