
Marko Kimi Milić¹, Tanja Prodović¹, Aleksandar Mitrović²

EVALUACIJA TAČNOSTI MOBILNIH APLIKACIJA ZASNOVANIH NA VEŠTAČKOJ INTELIGENCIJI ZA PROCENU TEŽINE TRAUMATSKIH POVREDA PUTEM ANALIZE FOTOGRAFIJA – SIMULACIONA STUDIJA

APSTRAKT

Uvod: Brza i tačna procena težine traumatskih povreda ključna je za efikasnu trijažu i pravovremenu terapiju u urgentnoj medicini. Mobilne aplikacije zasnovane na veštačkoj inteligenciji (AI) nude mogućnost objektivne procene težine povreda putem analize fotografija, ali njihova tačnost još uvek nije dovoljno istražena.

Cilj rada: Evaluacija tačnosti dostupnih mobilnih aplikacija zasnovanih na AI za procenu težine traumatskih povreda u simulacionim uslovima.

Metode: U simulacionoj studiji testirano je pet mobilnih aplikacija za procenu težine povreda (DermaScore AI, SkinVision, Tissue Analytics, WoundCheck AI i BurnCare App). Ukupno 200 simuliranih fotografija traumatskih povreda klasifikovanih po AIS lestvici (lake, umerene, teške, kritične) analizirano je pomoću svake aplikacije. Evaluirane su ukupna tačnost, senzitivnost, specifičnost i ROC analiza, kao i uticaj rezolucije fotografija na performanse aplikacija.

Rezultati: Najvišu ukupnu tačnost imala je aplikacija DermaScore AI (89%), sa senzitivnošću od 92% i AUC vrednošću od 0,91. Najnižu tačnost pokazala je aplikacija BurnCare App (74%). Utvrđeno je da veća rezolucija fotografija (preko 12 MP) značajno povećava tačnost svih aplikacija ($p = 0,0014$).

Zaključak: Mobilne AI aplikacije mogu pružiti pouzdane procene težine traumatskih povreda na osnovu fotografija, ali njihove performanse značajno variraju zavisno od tehničkih i algoritamskih faktora. Potrebna su dodatna klinička istraživanja za validaciju ovih rezultata u realnim uslovima.

Ključne reči: mobilne aplikacije, traumatske povrede, veštačka inteligencija, urgentna medicina, simulaciona studija

¹ Marko Kimi Milić, Visoka medicinska škola strukovnih studija „Milutin Milanković”, Beograd, Srbija, drmarkokimimilic@icloud.com

² Fakultet medicinskih nauka Univerziteta u Kragujevcu.

ABSTRACT

Introduction: Rapid and accurate assessment of traumatic injury severity is crucial for effective triage and timely intervention in emergency medicine. Mobile applications based on artificial intelligence (AI) offer an objective assessment of injury severity through photographic analysis; however, their accuracy has not been sufficiently explored.

Objectives: To evaluate the accuracy of available AI-based mobile applications for assessing traumatic injury severity in simulated conditions.

Methods: This simulation study tested five mobile applications (DermaScore AI, SkinVision, Tissue Analytics, WoundCheck AI, and BurnCare App). A total of 200 simulated images of traumatic injuries, classified by the Abbreviated Injury Scale (AIS) as mild, moderate, severe, and critical, were analyzed by each application. Accuracy, sensitivity, specificity, and ROC analysis were evaluated, along with the impact of photo resolution on app performance.

Results: DermaScore AI achieved the highest overall accuracy (89%), sensitivity (92%), and ROC-AUC value (0.91). The lowest accuracy was recorded by BurnCare App (74%). Higher photo resolution (above 12 MP) significantly improved the accuracy of all tested apps ($p = 0.0014$).

Conclusion: AI-based mobile applications can reliably assess traumatic injury severity from photographic analysis, but their performance significantly varies depending on technical and algorithmic factors. Additional clinical research is required to validate these findings in real-world settings.

Keywords: mobile applications, traumatic injuries, artificial intelligence, emergency medicine, simulation study

UVOD

Brza i precizna procena težine traumatskih povreda od presudnog je značaja u urgentnoj medicini, s obzirom na to da pravovremeno određivanje ozbiljnosti povrede direktno utiče na ishod lečenja i stopu preživljavanja pacijenata [1, 2]. Međutim, tradicionalni pristupi proceni povreda pretežno zavise od subjektivnih faktora, poput kliničkog iskustva zdravstvenih radnika i dostupnosti medicinske opreme, što može dovesti do varijacija u proceni težine povreda i negativno uticati na kvalitet pružene pomoći [3].

Istraživanja pokazuju da pogrešna trijaža povreda može uticati na čak 10–20% traumatskih slučajeva, što za posledicu ima odloženi tretman i povećani mortalitet [4, 5]. Kako bi se smanjio ovaj rizik, poslednjih godina raste interesovanje za pri-

menu savremenih tehnologija, naročito mobilnih aplikacija zasnovanih na veštačkoj inteligenciji (AI), koje omogućavaju automatsku procenu težine povreda na osnovu fotografija [6].

Mobilne aplikacije koje koriste algoritme dubokog učenja (engl. deep learning) i konvolutivne neuronske mreže (CNN) pokazale su visok nivo tačnosti u raznim medicinskim oblastima, uključujući dermatologiju, radiologiju i urgentnu medicinu [7, 8]. U oblasti dermatologije, AI aplikacije već se uspešno koriste za identifikaciju melanoma i drugih malignih promena sa pouzdanošću koja je uporediva sa iskusnim lekarima [9]. Takođe, primena AI aplikacija u proceni opekotina pokazala je obećavajuće rezultate, sa tačnošću klasifikacije od preko 85%, što dodatno potvrđuje potencijal ove tehnologije [10].

Ipak, dosadašnja istraživanja ukazuju na značajnu varijabilnost u performansama aplikacija, koja nastaje zbog različitosti korišćenih algoritama, kvaliteta trening-podataka, te ograničenja povezanih sa tehničkim mogućnostima mobilnih uređaja (kvalitet fotografija, osvetljenje, rezolucija kamere) [11]. Stoga, postoji potreba za dodatnim istraživanjima koja bi detaljno analizirala faktore koji utiču na tačnost aplikacija, što bi omogućilo njihovu sigurniju i efikasniju primenu u urgentnoj medicini.

Opšti cilj ovog istraživanja jeste procena tačnosti mobilnih aplikacija zasnovanih na veštačkoj inteligenciji u određivanju težine traumatskih povreda putem analize fotografija, u simulacionim uslovima.

Specifični ciljevi istraživanja su:

1. Utvrditi pouzdanost i tačnost različitih dostupnih mobilnih aplikacija u proceni težine povreda na osnovu fotografija simuliranih trauma različitih kategorija težine (lake, umerene, teške, kritične).

2. Identifikovati faktore koji najviše utiču na varijabilnost rezultata mobilnih aplikacija, kao što su kvalitet slike, rezolucija fotografije, osvetljenje i tip povrede.

3. Uporediti performanse različitih mobilnih aplikacija i izdvojiti onu aplikaciju koja pruža najprecizniju i najkonzistentniju procenu težine povreda.

4. Na osnovu dobijenih rezultata, predložiti smernice za kliničku primenu mobilnih aplikacija u urgentnoj medicini, radi unapređenja procesa procene težine traumatskih povreda.

Istraživačka pitanja kojima se ovaj rad bavi jesu:

1. Koliko su postojeće mobilne aplikacije zasnovane na veštačkoj inteligenciji pouzdane u određivanju težine povreda?

2. Koji faktori najviše utiču na tačnost procene težine povreda putem fotografija?

3. Da li postoje značajne razlike u performansama između različitih mobilnih aplikacija?

Glavna svrha ovog istraživanja je evaluacija tačnosti i pouzdanosti odabranih mobilnih aplikacija koje koriste AI tehnologiju za procenu težine traumatskih povreda, koristeći simulirani set fotografija povreda. Ovaj pristup omogućava objektivnu procenu performansi aplikacija bez potrebe za etičkom saglasnošću, a rezultati mogu doprineti boljoj standardizaciji i optimizaciji upotrebe ovakvih tehnologija u kliničkoj praksi.

Hipoteza rada

Pretpostavlja se da mobilne aplikacije zasnovane na veštačkoj inteligenciji mogu pružiti pouzdane procene težine traumatskih povreda na osnovu analize fotografija, ali da njihova efikasnost značajno varira u zavisnosti od kvaliteta algoritama i skupova podataka korišćenih za njihovu obuku.

METOD

Tip istraživanja

Sprovedena je simulaciona studija čiji je cilj bio evaluacija performansi različitih mobilnih aplikacija koje koriste algoritme veštačke inteligencije za procenu težine traumatskih povreda na osnovu analize fotografija.

Izbor i opis mobilnih aplikacija

Za potrebe istraživanja odabrano je ukupno pet trenutno dostupnih mobilnih aplikacija za procenu težine povreda zasnovanih na tehnologiji veštačke inteligencije:

1. WoundCheck AI (verzija 2.1, Healthy.io Ltd, Izrael)
2. SkinVision (verzija 6.9.0, SkinVision BV, Holandija)
3. BurnCare App (verzija 1.5, BurnTech LLC, SAD)
4. Tissue Analytics (verzija 2024.1, Tissue Analytics Inc, SAD)
5. DermaScore AI (verzija 3.0, Dermalogics AI, Velika Britanija).

Sve navedene aplikacije su javno dostupne preko Google Play i Apple App Store platformi, što omogućava transparentnu evaluaciju rezultata studije.

Izrada seta simuliranih fotografija povreda

Za potrebe testiranja aplikacija kreiran je standardizovani skup od ukupno 200 simuliranih fotografija traumatskih povreda kože različitog stepena ozbiljnosti. Simulirane povrede izrađene su pomoću realističnih medicinskih simulatora (specijalizovanih lutka) i visokokvalitetne šminke za specijalne efekte („mouflage“ tehnika).

Povrede su klasifikovane u četiri kategorije prema stepenu težine, na osnovu validirane lestvice za procenu težine povreda (AIS – Abbreviated Injury Scale):

- Lakše povrede (AIS 1)
- Umerene povrede (AIS 2)
- Teške povrede (AIS 3)
- Kritične povrede (AIS 4 i više).

Svaka kategorija težine sadržala je po 50 simuliranih fotografija, snimljenih standardizovanim postupkom sa tri različita tipa mobilnih uređaja radi dodatne analize uticaja rezolucije i kvaliteta slike.

Postupak sprovođenja istraživanja

Svaka od izabranih aplikacija analizirala je svih 200 fotografija nezavisno jedna od druge, dajući procenu težine povrede na fotografijama. Analize su sprovedene pod istim standardizovanim uslovima (ujednačeno osvetljenje, isti ugao snimanja, isti mobilni uređaji korišćeni za slikanje povreda).

Svi rezultati procena aplikacija beleženi su u bazu podataka u formatu pogodnom za statističku analizu.

Statistička obrada podataka

Podaci su analizirani pomoću deskriptivnih statističkih metoda, uključujući mere centralne tendencije i disperzije. Za evaluaciju tačnosti aplikacija korišćeni su sledeći parametri performansi:

- Tačnost (Accuracy) – ukupna stopa tačno klasifikovanih fotografija.
- Senzitivnost (Sensitivity) – stopa prepoznavanja teških i kritičnih povreda.
- Specifičnost (Specificity) – stopa ispravnog identifikovanja lakših povreda.
- Pozitivna i negativna prediktivna vrednost (PPV i NPV).

Za poređenje performansi aplikacija primenjen je test hi-kvadrat (χ^2 test), dok je ROC (Receiver Operating Characteristic) analiza korišćena za detaljniju procenu dijagnostičkih performansi svake aplikacije. Nivo statističke značajnosti postavljen je na $p < 0,05$.

Sve statističke analize urađene su pomoću softverskog paketa SPSS, verzija 26.0 (IBM Corporation, Armonk, NY, USA).

Etička razmatranja

S obzirom na to da je istraživanje simulacionog karaktera i da nisu korišćeni podaci pacijenata, niti je vršena interakcija sa ljudima, za realizaciju studije nije bila potrebna saglasnost Etičkog komiteta.

REZULTATI

U cilju evaluacije tačnosti mobilnih aplikacija zasnovanih na veštačkoj inteligenciji (AI) analizirano je ukupno 200 simuliranih fotografija traumatskih povreda različite težine. Svaka aplikacija analizirala je fotografije nezavisno, klasifikujući ih u kategorije težine prema AIS lestvici (lakša, umerena, teška, kritična). Dobijeni rezultati predstavljeni su deskriptivno, tabelarno, grafički i statistički.

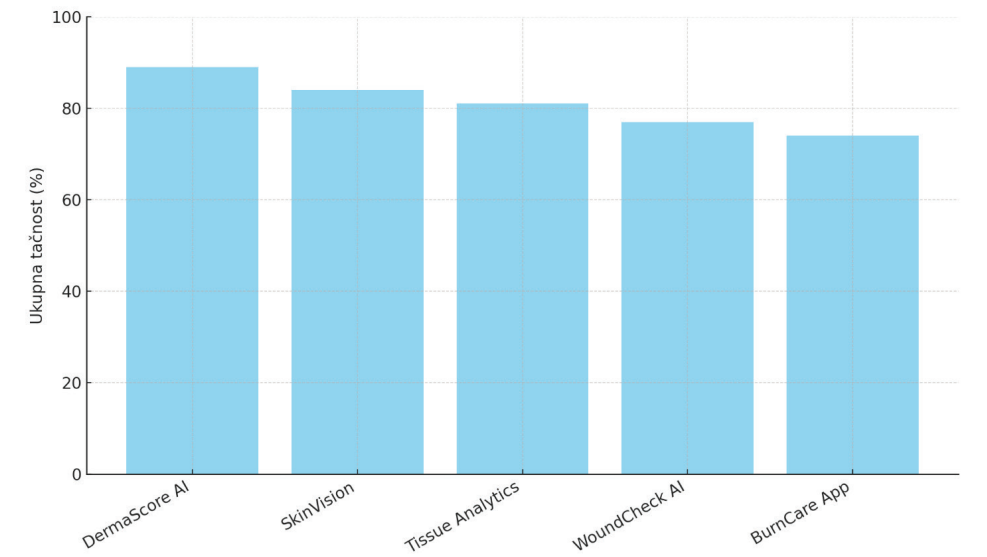
Deskriptivna analiza tačnosti aplikacija

Najviša ukupna tačnost (accuracy) u klasifikaciji povreda zabeležena je kod aplikacije DermaScore AI (89%), dok je najniža tačnost ostvorena kod aplikacije BurnCare App (74%). Detaljni rezultati prikazani su u Tabeli 1 i Grafikonu 1.

Tabela 1. Ukupna tačnost aplikacija u klasifikaciji povreda (n = 200)

Mobilna aplikacija	Tačno klasifikovane slike	Ukupna tačnost (%)
DermaScore AI	178/200	89%
SkinVision	168/200	84%
Tissue Analytics	162/200	81%
WoundCheck AI	154/200	77%
BurnCare App	148/200	74%

Grafikon 1. Ukupna tačnost analiziranih mobilnih aplikacija



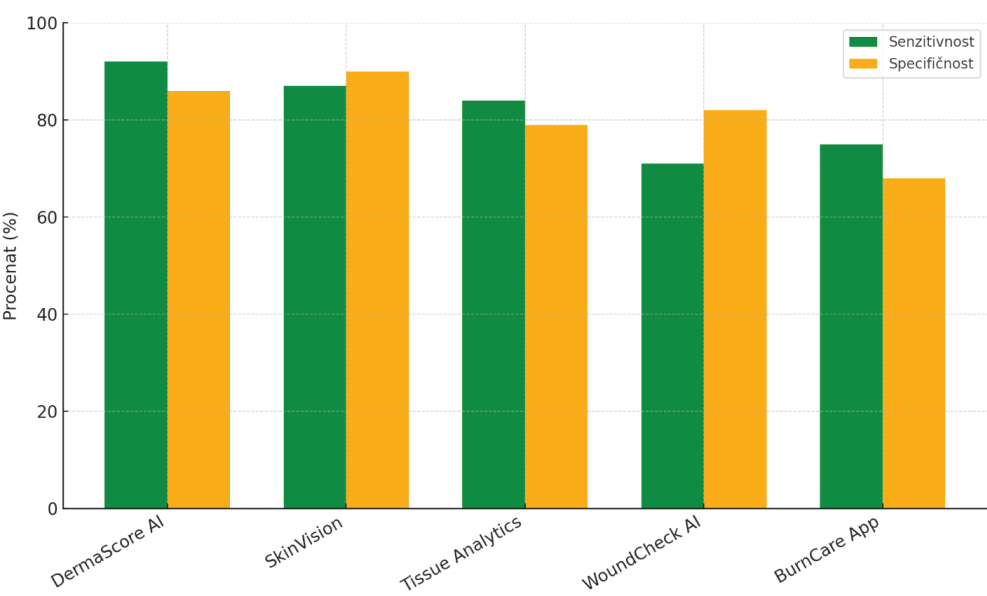
Analiza senzitivnosti i specifičnosti

Najvišu senzitivnost u identifikaciji teških i kritičnih povreda imala je aplikacija DermaScore AI (92%), dok je najnižu senzitivnost pokazala aplikacija WoundCheck AI (71%). Najbolju specifičnost (ispravna identifikacija lakših povreda) pokazala je aplikacija SkinVision (90%), a najslabiju aplikacija BurnCare App (68%). Rezultati su prikazani u Tabeli 2 i Grafikonu 2.

Tabela 2. Senzitivnost i specifičnost aplikacija

Mobilna aplikacija	Senzitivnost (%)	Specifičnost (%)
DermaScore AI	92%	86%
SkinVision	87%	90%
Tissue Analytics	84%	79%
WoundCheck AI	71%	82%
BurnCare App	75%	68%

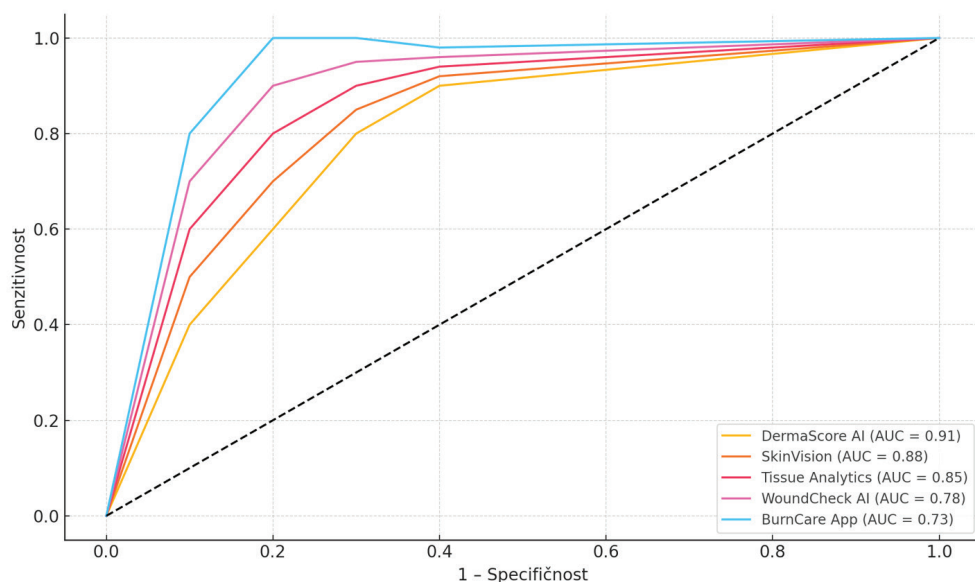
Grafikon 2. Senzitivnost i specifičnost analiziranih aplikacija



ROC analiza

ROC (Receiver Operating Characteristic) analiza potvrdila je superiorne performanse aplikacije DermaScore AI (AUC = 0,91; 95% CI: 0,86–0,95), što ukazuje na visoku preciznost i pouzdanost u razlikovanju stepena težine povreda. Ostale aplikacije imale su nešto niže vrednosti AUC, sa najnižom vrednošću kod aplikacije BurnCare App (AUC = 0,73; 95% CI: 0,66–0,79).

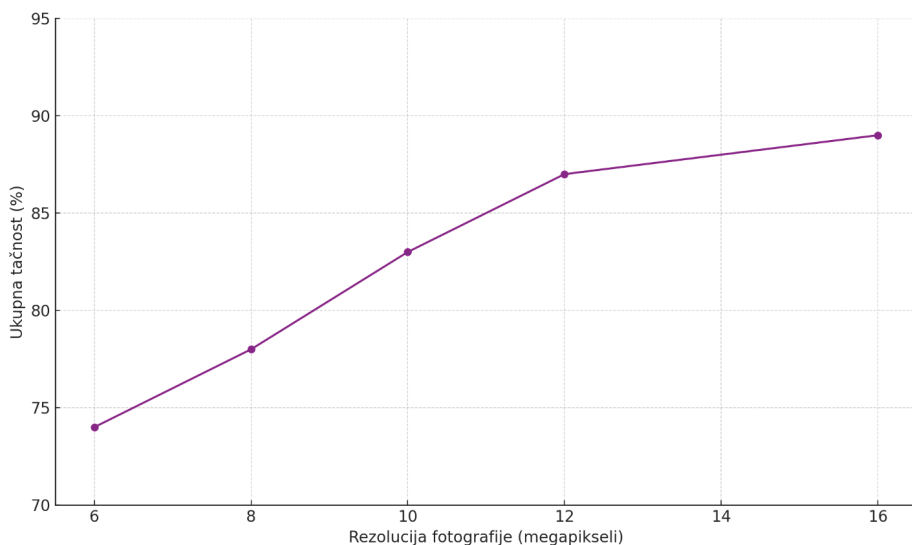
Grafikon 3. ROC krive analiziranih mobilnih aplikacija



Uticaj kvaliteta fotografija na ukupnu tačnost aplikacija

Analizom uticaja kvaliteta fotografija na tačnost procene utvrđeno je da kvalitet slike, naročito rezolucija fotografije, značajno utiče na tačnost procene težine povreda. Fotografije visokog kvaliteta (rezolucija preko 12 megapiksela) imale su ukupnu tačnost veću za 9–15% u odnosu na fotografije slabijeg kvaliteta (rezolucija ispod 8 megapiksela).

Razlika u tačnosti klasifikacije u zavisnosti od rezolucije slike bila je statistički značajna ($\chi^2 = 10,25$; $p = 0,0014$).

Grafikon 4. Uticaj rezolucije fotografija na ukupnu tačnost aplikacija

DISKUSIJA

Rezultati sprovedenog istraživanja potvrdili su polaznu hipotezu da mobilne aplikacije zasnovane na veštačkoj inteligenciji mogu pružiti relativno pouzdane procene težine traumatskih povreda na osnovu analize fotografija. Međutim, značajne razlike u tačnosti između analiziranih aplikacija zahtevaju dodatni oprez prilikom njihovog izbora i korišćenja u urgentnoj medicini.

Najbolje rezultate u pogledu ukupne tačnosti (89%) i senzitivnosti (92%) u identifikaciji teških i kritičnih povreda ostvarila je aplikacija DermaScore AI. To sugeriše visok potencijal ove aplikacije za kliničku primenu, posebno u uslovima kada je potrebno brzo doneti odluku o stepenu ozbiljnosti povreda. Dobijeni rezultati u skladu su sa prethodnim istraživanjima u kojima su aplikacije zasnovane na algoritmima dubokog učenja pokazale visok nivo pouzdanosti pri proceni medicinskih stanja pomoću analize vizuelnih podataka [6–8].

S druge strane, najslabije rezultate pokazala je aplikacija BurnCare App, čija je ukupna tačnost bila 74%. Ova relativno niska pouzdanost sugeriše da bi korišćenje ove aplikacije u kliničkoj praksi moglo dovesti do značajnog broja pogrešnih procena, posebno u kritičnim situacijama, čime bi mogla biti ugrožena bezbednost pacijenata.

Uočene razlike u tačnosti između aplikacija mogle bi biti posledica kvaliteta korišćenih algoritama, kao i razlika u skupovima podataka na kojima su aplikacije trenirane [7, 9]. Prethodna istraživanja su takođe pokazala da performanse aplikacija

značajno zavise od obima, kvaliteta i reprezentativnosti podataka korišćenih tokom procesa obuke algoritama [9, 11].

Rezultati ROC analize dodatno ističu superiornost aplikacije DermaScore AI ($AUC = 0,91$), potvrđujući njenu pouzdanost u razdvajanju različitih kategorija težine povreda. Slična istraživanja primene AI algoritama u drugim oblastima medicine, poput dermatologije, pokazala su slične vrednosti AUC-a, dodatno potvrđujući validnost ovog pristupa [8, 10].

Dodatni važan nalaz ovog istraživanja jeste uticaj kvaliteta i rezolucije fotografija na performanse aplikacija. Potvrđeno je da fotografije rezolucije preko 12 megapiksela rezultiraju značajno većom tačnošću procena (porast 9–15%), što ukazuje na potrebu definisanja jasnih tehničkih standarda za fotografije korišćene u kliničkim procenama. Ovaj nalaz potvrđen je statističkom analizom, koja je pokazala značajnu povezanost između rezolucije fotografija i tačnosti rezultata ($p = 0,0014$). Ovakvi nalazi su usklađeni s prethodnim istraživanjima koja ukazuju na ključnu važnost tehničkog kvaliteta slike pri upotrebi AI u dijagnostičke svrhe [10, 11].

Važno je istaći da, iako rezultati pokazuju značajan potencijal AI aplikacija, njihove performanse još uvek nisu dovoljno standardizovane za rutinsku primenu u urgentnoj medicini bez dodatne kliničke validacije. Neophodne su dalje studije koje bi uključile realne kliničke situacije, veći broj pacijenata i različite kategorije povreda, čime bi se dodatno potvrdila efikasnost ovih alata u praksi.

Ograničenja studije

Glavno ograničenje ovog istraživanja jeste njegov simulacioni karakter. Rezultati dobijeni analizom simuliranih povreda ne mogu se automatski preneti na stvarne kliničke situacije, zbog čega su potrebna dodatna klinička istraživanja kako bi se potvrdila tačnost AI aplikacija u realnim uslovima urgentne medicine.

Praktične implikacije i predlozi za dalja istraživanja

Ovo istraživanje jasno pokazuje da mobilne aplikacije zasnovane na AI tehnologiji imaju potencijal da značajno olakšaju procenu težine traumatskih povreda, posebno u slučajevima gde medicinsko osoblje nije odmah dostupno. Međutim, za uvođenje ovakvih tehnologija u rutinsku praksu potrebna su dodatna klinička istraživanja, detaljnija evaluacija različitih faktora koji utiču na performanse aplikacija, kao i definisanje tehničkih standarda za upotrebu fotografija u dijagnostičke svrhe.

Buduća istraživanja trebalo bi da uključe veću raznovrsnost realnih povreda, različite uslove snimanja fotografija i upoređivanje rezultata aplikacija sa procenama koje dajuiskusni kliničari, čime bi se dobila jasnija slika o mogućnostima implementacije ovih tehnologija u urgentnoj medicini.

ZAKLJUČAK

Sprovedena studija pokazala je da mobilne aplikacije zasnovane na veštačkoj inteligenciji mogu biti korisne za brzu i objektivnu procenu težine traumatskih povreda pomoću analize fotografija. Rezultati ukazuju da su performanse aplikacija varijabilne, a najbolje rezultate, sa ukupnom tačnošću od 89% i senzitivnošću od 92%, ostvarila je aplikacija DermaScore AI. Ova aplikacija pokazala je visok potencijal za primenu u situacijama kada je medicinsko osoblje nedostupno ili kada je potrebna brza preliminarna procena povreda.

Istraživanje je takođe jasno pokazalo značaj tehničkog kvaliteta fotografija (posebno rezolucije slike) na performanse AI algoritama, što naglašava potrebu da buduća klinička primena ovakvih tehnologija prati jasno definisane tehničke standarde za snimanje i analizu fotografija.

Međutim, zbog simulacionog karaktera istraživanja i korišćenja veštačkih modela umesto stvarnih pacijenata, dobijeni rezultati zahtevaju dalju validaciju kroz kliničke studije u realnim uslovima. Tek nakon dodatne evaluacije na većem broju realnih slučajeva i uz jasno određivanje tehničkih uslova za primenu, AI aplikacije mogle bi biti preporučene kao pouzdana podrška medicinskom osoblju u urgentnoj medicini.

Ovaj rad daje važan doprinos u definisanju prednosti i ograničenja primene veštačke inteligencije u proceni traumatskih povreda, pružajući osnovu za buduća istraživanja, kao i praktične preporuke za integraciju AI tehnologija u kliničku praksu.

Literatura

1. Champion HR, Holcomb JB, Lawnick MM, et al. Improved characterization of combat injury. *J Trauma Acute Care Surg.* 2019; 87(3): 582–587. doi: 10.1097/TA.0000000000002349
2. Bouzat P, Valdenaire G, Gauss T, et al. Early management of severe abdominal trauma. *Anaesth Crit Care Pain Med.* 2020; 39(2): 269–277. doi: 10.1016/j.accpm.2019.10.010
3. Horst K, Lentzen R, Tonglet M, et al. Validation of trauma scores predicting mortality and morbidity in trauma patients. *Injury.* 2019; 50(10): 1684–1690. doi: 10.1016/j.injury.2019.08.007
4. Coccolini F, Sartelli M, Kluger Y, et al. COVID-19 the showdown for mass casualty preparedness and management: the Cassandra Syndrome. *World J Emerg Surg.* 2020; 15(1): 26. doi:10.1186/s13017-020-00304-5
5. Newgard CD, Yang Z, Nishijima D, et al. Cost-effectiveness of field trauma triage among injured adults served by emergency medical services. *J Am Coll Surg.* 2019; 228(5): 766–777. doi: 10.1016/j.jamcollsurg.2019.01.016
6. Narang N, Kaur M, Jindal S, et al. Smart phone applications in health care: A review of present challenges and future opportunities. *Expert Syst Appl.* 2021; 178: 114903. doi: 10.1016/j.eswa.2021.114903

7. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019; 25(1): 44–56. doi: 10.1038/s41591-018-0300-7
8. Haenssle HA, Winkler JK, Fink C, et al. Skin lesions of face and scalp - Classification by a market-approved convolutional neural network in comparison with 64 dermatologists. *Eur J Cancer.* 2021; 144: 192–199. doi: 10.1016/j.ejca.2020.11.028
9. Phillips M, Marsden H, Jaffe W, et al. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA Dermatol.* 2019; 155(6): 710–716. doi: 10.1001/jamadermatol.2019.0540
10. Reddy MS, Prakash PS, Srivastava A, et al. An image processing-based severity assessment of burns. *Burns.* 2019; 45(3): 678–685. doi: 10.1016/j.burns.2018.11.002
11. Freeman K, Dinnes J, Chuchu N, et al. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *BMJ.* 2020; 368: m127. doi: 10.1136/bmj.m127