Marko Kimi Milić[1], Tanja Prodović[1], Aleksandar Mitrović[2]

ACCURACY EVALUATION OF ARTIFICIAL INTELLIGENCE-BASED MOBILE APPLICATIONS FOR ASSESSING THE SEVERITY OF TRAUMATIC INJURIES THROUGH PHOTO ANALYSIS – A SIMULATION STUDY

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

[1] Marko Kimi Milić, High Medical College of Professional Studies “Milutin Milankovic”, Belgrade, Serbia, drmarkokimimilic@icloud.com

[2] Faculty of Medical Sciences, University of Kragujevac

INTRODUCTION

Rapid and accurate assessment of the severity of traumatic injuries is of critical importance in emergency medicine, as timely determination of injury seriousness directly influences treatment outcomes and patient survival rates [1,2]. However, traditional approaches to injury assessment largely depend on subjective factors such as the clinical experience of healthcare professionals and the availability of medical equipment, which can lead to inconsistencies in injury severity evaluation and negatively affect the quality of care provided [3].

Studies have shown that incorrect injury triage may affect as many as 10–20% of trauma cases, resulting in delayed treatment and increased mortality [4,5]. To mitigate this risk, there has been growing interest in recent years in the use of modern technologies, particularly mobile applications based on artificial intelligence (AI), which enable automated injury severity assessment using photographs [6].

Mobile applications that utilize deep learning algorithms and convolutional neural networks (CNNs) have demonstrated high accuracy in various medical fields, including dermatology, radiology, and emergency medicine [7,8]. In dermatology, AI-based applications are already successfully used for melanoma detection and the identification of other malignant lesions, with reliability comparable to that of experienced physicians [9]. Moreover, the use of AI applications for burn assessment has shown promising results, with classification accuracy exceeding 85%, further confirming the potential of this technology [10].

Nevertheless, current studies indicate significant variability in application performance, stemming from differences in algorithms, the quality of training datasets, and limitations related to the technical capabilities of mobile devices (image quality, lighting conditions, camera resolution) [11]. Therefore, further research is needed to analyze the factors influencing application accuracy, which would enable their safer and more effective use in emergency medical settings.

The general aim of this study is to assess the accuracy of AI-based mobile applications in determining the severity of traumatic injuries through photographic analysis, under simulated conditions.

Specific objectives of the study include:

1. To determine the reliability and accuracy of various available mobile applications in assessing injury severity based on photographs of simulated traumas of different severity levels (mild, moderate, severe, critical).

2. To identify the key factors that influence the variability in mobile application results, such as image quality, photo resolution, lighting, and type of injury.

3. To compare the performance of different mobile applications and highlight the one that provides the most accurate and consistent injury severity assessment.

4. Based on the obtained results, to propose guidelines for clinical application of mobile applications in emergency medicine, with the aim of improving the injury severity assessment process.

The research questions addressed in this paper are:

1. How reliable are current AI-based mobile applications in determining injury severity?

2. What are the key factors influencing the accuracy of injury severity assessment using photographs?

3. Are there significant differences in performance among various mobile applications?

The main purpose of this study is to evaluate the accuracy and reliability of selected mobile applications that utilize AI technology for assessing the severity of traumatic injuries, using a simulated dataset of injury photographs. This approach allows for an objective evaluation of app performance without the need for ethical approval, and the results can contribute to better standardization and optimization of such technologies in clinical practice.

Study Hypothesis

It is hypothesized that AI-based mobile applications can provide reliable assessments of traumatic injury severity based on photo analysis, but their effectiveness significantly varies depending on the quality of the algorithms and the datasets used during training.

METHOD

Type of Study

A simulation study was conducted with the aim of evaluating the performance of various mobile applications that utilize artificial intelligence algorithms to assess the severity of traumatic injuries based on photographic analysis.

Selection and Description of Mobile Applications

For the purposes of this study, five currently available AI-based mobile applications for injury severity assessment were selected:

1. WoundCheck AI (version 2.1, Healthy.io Ltd, Israel)

2. SkinVision (version 6.9.0, SkinVision BV, Netherlands)

3. BurnCare App (version 1.5, BurnTech LLC, USA)

4. Tissue Analytics (version 2024.1, Tissue Analytics Inc, USA)

5. DermaScore AI (version 3.0, Dermalogics AI, United Kingdom)

All listed applications are publicly available through Google Play and Apple App Store platforms, allowing for transparent evaluation of study results.

Creation of Simulated Injury Photo Set

To test the applications, a standardized set of 200 simulated photographs of skin injuries of varying severity was created. The simulated injuries were produced using realistic medical manikins and high-quality special effects makeup (moulage technique).

Injuries were classified into four categories based on the validated Abbreviated Injury Scale (AIS):

- Mild injuries (AIS 1)
- Moderate injuries (AIS 2)
- Severe injuries (AIS 3)
- Critical injuries (AIS 4 and above)

Each severity category included 50 simulated photographs, captured using a standardized procedure with three different types of mobile devices to further analyze the impact of resolution and image quality.

Study Procedure

Each of the selected applications analyzed all 200 photographs independently, providing a severity assessment for each image. All analyses were conducted under the same standardized conditions (uniform lighting, identical camera angles, and the same mobile devices used for capturing images).

All assessment results from the applications were recorded in a database suitable for statistical analysis.

Statistical Data Analysis

Data were analyzed using descriptive statistical methods, including measures of central tendency and dispersion. The following performance metrics were used to evaluate application accuracy:

- Accuracy – overall rate of correctly classified images.
- Sensitivity – rate of correctly identified severe and critical injuries.

- Specificity – rate of correctly identified mild injuries.
- Positive and Negative Predictive Values (PPV and NPV).

To compare application performance, the chi-square test (χ^2 test) was used, and ROC (Receiver Operating Characteristic) analysis was applied for a more detailed evaluation of the diagnostic performance of each application. The level of statistical significance was set at $p < 0.05$.

All statistical analyses were performed using SPSS software, version 26.0 (IBM Corporation, Armonk, NY, USA).

Ethical Considerations

Since this was a simulation study and did not involve patient data or human interaction, ethical committee approval was not required for the execution of this research.

RESULTS

To evaluate the accuracy of artificial intelligence (AI)-based mobile applications, a total of 200 simulated photographs of traumatic injuries of varying severity were analyzed. Each application independently assessed the photographs, classifying them into severity categories based on the AIS (Abbreviated Injury Scale): mild, moderate, severe, and critical. The obtained results are presented descriptively, in tabular form, graphically, and statistically.

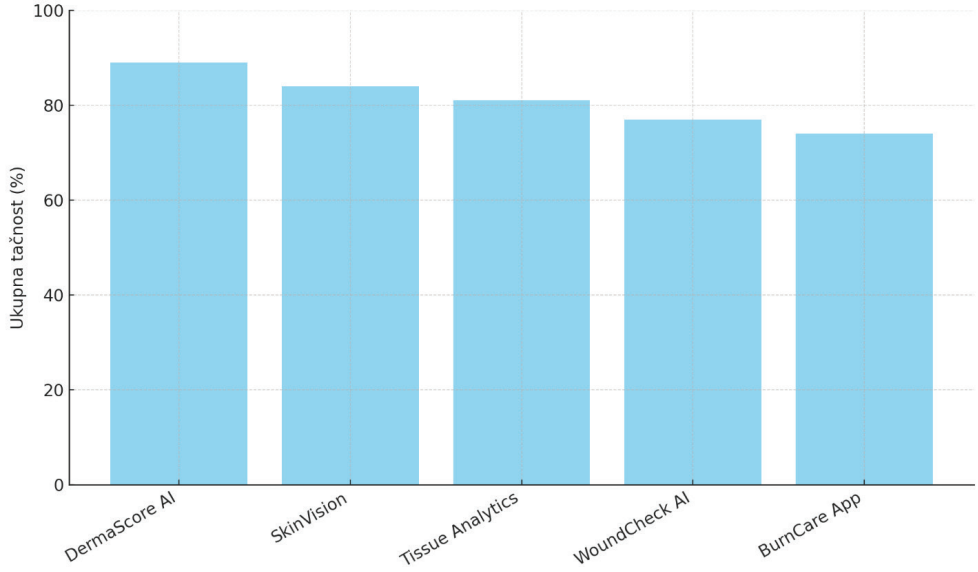
Descriptive Analysis of Application Accuracy

The highest overall accuracy in injury classification was recorded for the DermaScore AI application (89%), while the lowest accuracy was observed in the BurnCare App (74%). Detailed results are shown in Table 1 and Figure 1.

Table 1. Overall accuracy of applications in injury classification (n=200)

Mobile Application	Correctly Classified Images	Overall Accuracy (%)
DermaScore AI	178/200	89%
SkinVision	168/200	84%
Tissue Analytics	162/200	81%
WoundCheck AI	154/200	77%
BurnCare App	148/200	74%

Figure 1. Overall Accuracy of the Analyzed Mobile Applications



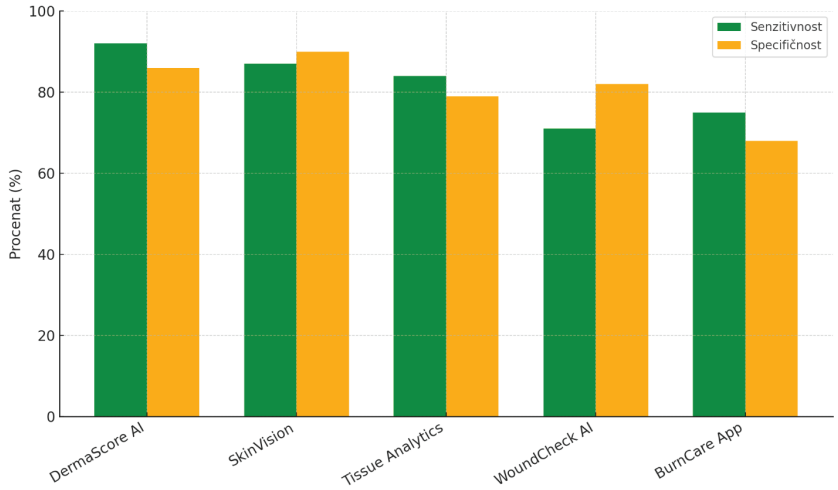
Sensitivity and Specificity Analysis

The highest sensitivity in identifying severe and critical injuries was observed in the DermaScore AI application (92%), while the lowest sensitivity was shown by WoundCheck AI (71%). The best specificity (accurate identification of mild injuries) was achieved by SkinVision (90%), whereas the lowest was recorded for BurnCare App (68%). The results are presented in Table 2 and Figure 2.

Table 2. Sensitivity and Specificity of Applications

Mobile Application	Sensitivity (%)	Specificity (%)
DermaScore AI	92%	86%
SkinVision	87%	90%
Tissue Analytics	84%	79%
WoundCheck AI	71%	82%
BurnCare App	75%	68%

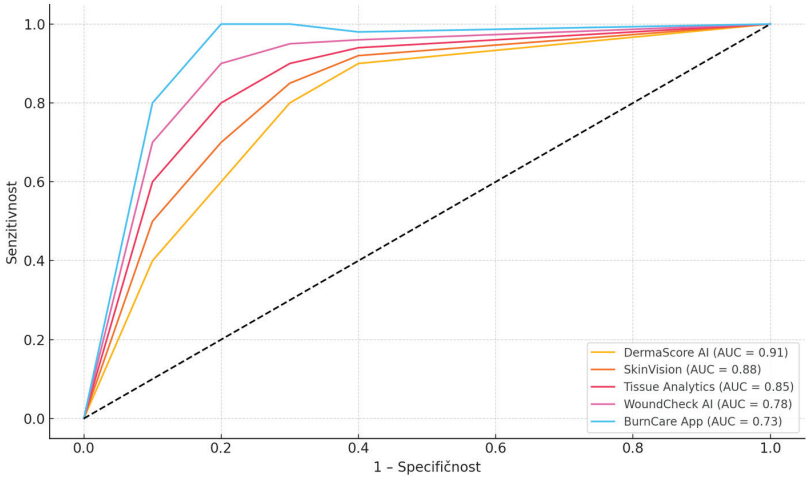
Figure 2. Sensitivity and Specificity of the Analyzed Applications



ROC Analysis

The ROC (Receiver Operating Characteristic) analysis confirmed the superior performance of the DermaScore AI application (AUC = 0.91; 95% CI: 0.86–0.95), indicating high precision and reliability in distinguishing injury severity levels. The other applications demonstrated slightly lower AUC values, with the lowest recorded for the BurnCare App (AUC = 0.73; 95% CI: 0.66–0.79).

Figure 3. ROC Curves of the Analyzed Mobile Applications

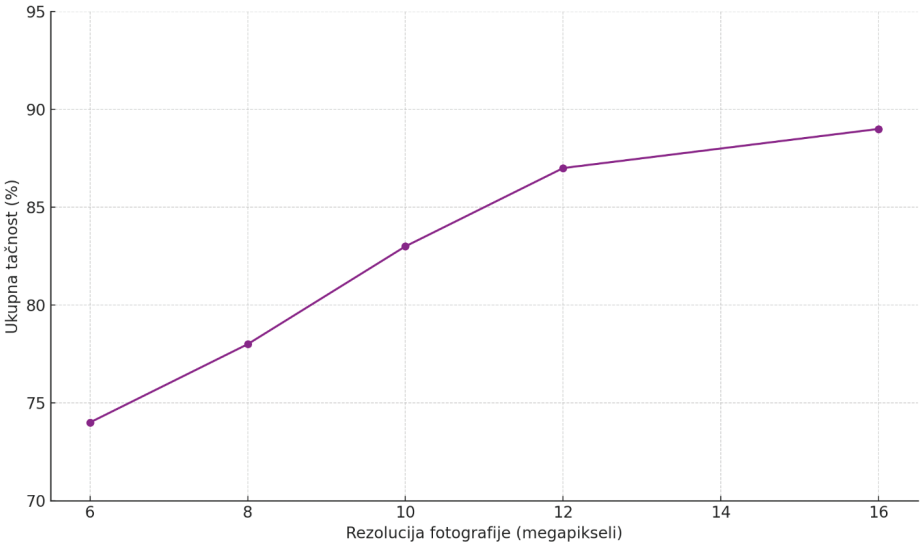


Impact of Photo Quality on Overall Application Accuracy

Analysis of the impact of photo quality on assessment accuracy revealed that image quality—particularly photo resolution—significantly affects the accuracy of injury severity evaluations. High-quality images (resolution over 12 megapixels) demonstrated an overall accuracy increase of 9–15% compared to lower-quality images (resolution below 8 megapixels).

The difference in classification accuracy based on image resolution was statistically significant ($\chi^2 = 10.25$; $p = 0.0014$).

Figure 4. Impact of Photo Resolution on Overall Application Accuracy



DISCUSSION

The results of this study confirmed the initial hypothesis that mobile applications based on artificial intelligence can provide relatively reliable assessments of traumatic injury severity through photo analysis. However, the significant differences in accuracy among the analyzed applications call for caution when selecting and using these tools in emergency medicine.

The best results in terms of overall accuracy (89%) and sensitivity (92%) for identifying severe and critical injuries were achieved by the DermaScore AI application. This suggests strong potential for its clinical use, especially in situations where rapid decision-making regarding injury severity is required. These results are consistent with

previous studies where deep learning-based applications demonstrated high reliability in medical condition assessment using visual data analysis [6–8].

In contrast, the weakest performance was recorded by the BurnCare App, with an overall accuracy of 74%. This relatively low reliability implies that the use of this app in clinical practice could lead to a considerable number of misclassifications, particularly in critical situations, thereby potentially compromising patient safety.

The observed differences in accuracy across applications could be attributed to the quality of the algorithms used, as well as variations in the training datasets on which the applications were developed [7,9]. Previous research has also shown that application performance is highly dependent on the scope, quality, and representativeness of the data used during the algorithm training process [9,11].

The ROC analysis further highlights the superiority of the DermaScore AI application (AUC = 0.91), reinforcing its reliability in distinguishing between different injury severity categories. Similar studies applying AI algorithms in other areas of medicine, such as dermatology, have reported comparable AUC values, further validating this approach [8,10].

Another important finding of this study is the impact of image quality and resolution on application performance. It was confirmed that photographs with a resolution above 12 megapixels resulted in significantly higher assessment accuracy (an increase of 9–15%), emphasizing the need to define clear technical standards for photographs used in clinical evaluations. This finding was supported by statistical analysis, which showed a significant correlation between photo resolution and assessment accuracy ($p = 0.0014$). These results are in line with previous studies that highlight the critical importance of technical image quality in the use of AI for diagnostic purposes [10,11].

It is important to emphasize that, although the results demonstrate the potential of AI applications, their performance is not yet sufficiently standardized for routine use in emergency medicine without additional clinical validation. Further studies are required involving real clinical scenarios, a larger number of patients, and different injury categories to further confirm the effectiveness of these tools in practice.

Study Limitations

The main limitation of this research is its simulation-based design. The results obtained from analyzing simulated injuries cannot be directly extrapolated to real clinical scenarios, which is why additional clinical research is necessary to validate the accuracy of AI applications under real-world emergency conditions.

Practical Implications and Suggestions for Future Research

This study clearly shows that AI-based mobile applications have the potential to significantly facilitate the assessment of traumatic injury severity, particularly in

situations where medical personnel are not immediately available. However, introducing such technologies into routine practice requires further clinical studies, more detailed evaluation of the various factors influencing application performance, and the definition of technical standards for the use of photographs in diagnostic contexts.

Future research should include a broader variety of real injuries, diverse photo-capturing conditions, and a comparison between application-generated assessments and evaluations provided by experienced clinicians, in order to gain a clearer understanding of the potential for implementing these technologies in emergency medicine.

CONCLUSION

This study demonstrated that mobile applications based on artificial intelligence can be useful tools for rapid and objective assessment of traumatic injury severity through photographic analysis. The results indicate that application performance varies, with the best results—an overall accuracy of 89% and a sensitivity of 92%—achieved by the DermaScore AI application. This app showed strong potential for use in situations where medical personnel are unavailable or when a quick preliminary injury assessment is needed.

The research also clearly highlighted the importance of the technical quality of images—particularly photo resolution—on the performance of AI algorithms, emphasizing the need for well-defined technical standards for image capture and analysis in future clinical applications.

However, due to the simulation-based nature of the study and the use of artificial models instead of real patients, the results require further validation through clinical studies in real-world settings. Only after additional evaluation involving a larger number of real cases and clearly defined technical conditions for implementation can AI applications be recommended as reliable support tools for medical staff in emergency medicine.

This study makes a valuable contribution to defining the advantages and limitations of using artificial intelligence in trauma assessment, providing a foundation for future research as well as practical recommendations for integrating AI technologies into clinical practice.

Reference

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