Real-time AI Image Classification System for Burn Injuries

Pei-Lun Sun[1], Meng-Yun Tsai[1*], Hsuan Su[1], Kuan-Ya Chen[2], Hsin-Yu Liu[1] , Ko-Hsin Hsiung[1] , Tzer-Long Chen[1*]

[1]Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University, Taiwan

[2]Department of Psychology, Kaohsiung Medical University, Taiwan

*Corresponding Author: u110029054@gap.kmu.edu.tw; tlchen@kmu.edu.tw

DOI:10.30211/JIC.202402.001

ABSTRACT

With the development of Taiwan's economy and public health education, burn injuries are no longer a leading cause of death in the country. Nevertheless, burn injuries are still unavoidable in our life. Thus, we must not underestimate the possibility of harmful consequences. In the past, large-scale deep learning models were employed to quickly classify the severity of burn injuries. Despite this, the recognition process still requires individuals to make a self-assessment regarding the need for medical attention. However, for users without professional medical knowledge, performing these tasks can be challenging. Additionally, given the widespread use of the Internet, as most users are accustomed to accessing information via mobile devices, expecting injured users to operate a desktop computer for wound classification is unrealistic. Therefore, this study aims to develop a web-based platform for classifying burn injuries that aligns with users' current habits. Users will be able to import wound images, and the system will automatically provide information regarding the severity of the burn injury, along with relevant medical advice. As for the classification system, we selected the GTM platform to train our deep learning model for optimal integration with the web server. Finally, we have rebuilt the training image dataset, which now includes 1,162 images of burned skin and 305 images of normal skin. The precision rates can achieve over 81% to 95% for all four degrees of burn injuries during the later stages of model verification. The customer experience with the classification platform also surpasses satisfactory levels.

Keywords: Burn injury, Artificial intelligence, Image classification, Grading burn injuries.

1. Introduction

Burn injuries are one of the most common types of trauma in people's daily lives. It is estimated that the total number of burn cases worldwide exceeds 9 million per year, resulting in an annual economic loss of more than 112 billion US dollars [1]. Although burns have a huge impact on society, Barclay et al (2022) pointed out that like most other traumas, the incidence and severity of burns can be reduced through preventive and medical measures, which is one of the main reasons for the generally lower incidence of burns in high-income countries [2]. However, the information available online for the public to conduct self-assessments is too complex and difficult to understand. There are also no tools available to the public that can quickly provide assessments of burn severity. This results in the public being unable to determine the need for medical care in the initial or follow-up observations, delaying medical interventions.

In recent years, artificial intelligence has been a very popular research topic in information technology. This is because, compared to the human brain, it can be faster and more effective in completing the same tasks [3]. As a result, the number of artificial intelligence research cases in the medical field has increased year by year, and the purpose has gradually extended from assisting the back-end of medicine to initial diagnosis at the front end. Artificial intelligence has already been widely used in trauma areas such as wound recognition. However, there are a few actual applications in the field of burns [4].

2. Literature Review

2.1 Image Dataset

This study uses the open academic research dataset from Kaggle as the research basis [5], and manually reviews the dataset to eliminate interfering images as much as possible and supplement missing feature images in the original dataset.

2.1.1 Image elimination

Before model training, we inspected the images in the original dataset. If the image contained features that could interfere with recognition, such as: tattoos on the body, markings covering the burned area, the affected area differing significantly from the actual situation, etc.; or the image contained identifiable personal information, such as: patient name, patient medical record number, etc., the image elimination operation was performed. This is to ensure the accuracy of the model after training, and to be as close to the actual application scenario as possible. All images are re-encoded after elimination to achieve de-identification.

2.1.2 Supplementation of wound feature images

In the preliminary verification of this study, we found that the model's ability to recognize second-degree burns was weaker. After checking the dataset, we found that some feature images of this classification were insufficient, such as blisters. To this end, we supplemented potential missing sample features for each burn classification to increase dataset diversity and improve the model's ability to identify each grade of burn.

2.2 Manual image grading labeling

The original image dataset already contained YOLO image grading labeling files provided by Kaggle. However, some image labeling files had missing or incorrect grading results. Therefore, we used Python to write a manual image review program. The program was used to confirm the grading of the labeled images or to add grading labels to photos with missing labels. After re-grading, the images of each grade finally used in this study are shown in Figure 1.

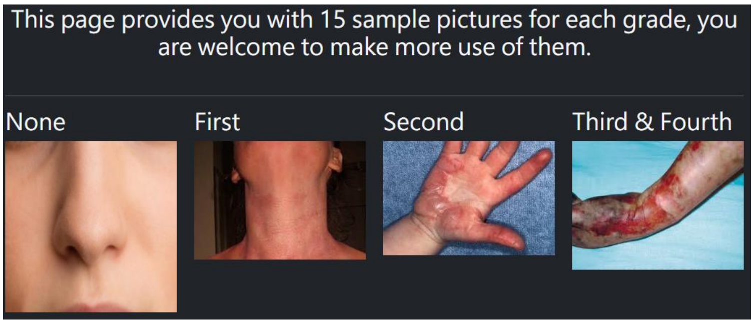


Fig. 1 Example images of different grading levels

2.3 Model Training

This study will use the Google Teachable Machine (GTM) platform to train and export the image recognition model. Thanks to the MobileNetV3 architecture foundation of GTM, the final exported model will be able to perform recognition tasks well on mobile devices, which will also help us deploy recognition models on servers.

2.3.1 Google teachable machine

GTM is a machine learning development platform provided by the Google team. The model used by GTM is the MobileNetV3 model developed by Google's team based on the Convolutional Neural Network (CNN) architecture of deep learning. Currently, GTM provides users with three types of projects to train models – image, audio and posture capture projects. The emergence of GTM allows students and enthusiasts without machine learning foundations to train and use machine learning through simplified training processes (Fig. 2). It is also because of the ease of use of this tool that promotes the popularization and application development of machine learning. All training image data in this study will be uploaded through the GTM platform and trained by Google Cloud Computing. Table 1 shows the final model training parameters used in this study.

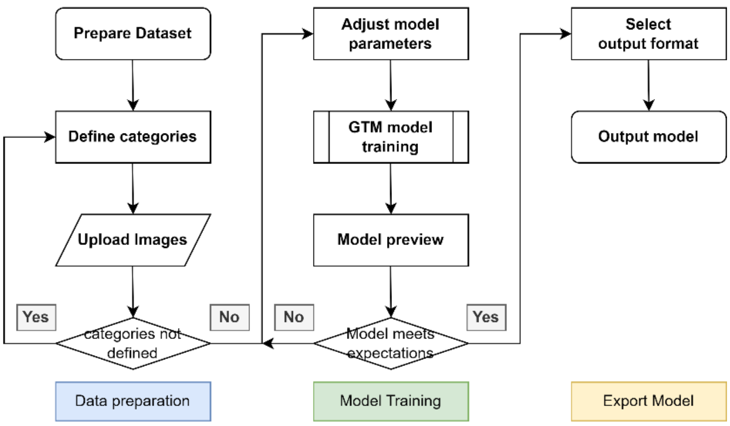


Fig. 2 GTM training process (image project as an example)

Table 1 Model training parameter settings

Param	Epochs	Batch Size	Learning Rate
Value	200	16	0.01

2.4 Model Effectiveness Measurement

We use three metrics to measure the effectiveness of the model trained in this study in predicting each classification. The formulas for the metrics can be calculated using the number of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) images [6].

2.4.1 Recall

The recall metric is defined as the proportion of samples in the test set of this classification that are correctly identified as belonging to that classification by the model. The calculation formula is shown in equation (1).

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \qquad (1)$$

2.4.2 Precision

The precision metric is defined as the proportion of samples that actually belong to that classification among all the samples identified by the model as that classification. The calculation formula is shown in equation (2).

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \qquad (2)$$

2.4.3 F1-Score

The F1-Score is a harmonic average mainly used to comprehensively measure the results of recall and precision, roughly evaluating the recognition performance of the model [7]. The calculation formula is shown in equation (3).

$$\mathrm{F1-Score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}, \qquad (3)$$

3. Research Design

3.1 Grading Recognition Effectiveness

Finally, after re-checking and supplementing samples for each grading in the dataset, a total of 1,467 external skin images were included in the training dataset, including 305 normal and 508 first degree, 461 second degree, 193 third and fourth degree burn images. Subsequently, the training dataset was randomly divided into training data input to GTM and test data for subsequent performance analysis at a ratio of 9:1 [8]. Table 2 shows the final recognition effectiveness of the model for each classification. Among them, the highest recall rate appeared in the third and fourth grades, at 0.9091; the highest prediction accuracy rate occurred in the non-burn classification, at 0.9533; In terms of comprehensive effectiveness, the recognition effectiveness of non-burns was the best. The recognition results provided by the system are shown in Fig. 3.

Table 2 Model grading result effectiveness table

Degree	F1-Score	Recall	Precision
None	0.9284	0.9048	0.9533
First	0.8149	0.8125	0.8173
Second	0.8485	0.8235	0.8751
Third & Fourth	0.8696	0.9091	0.8334

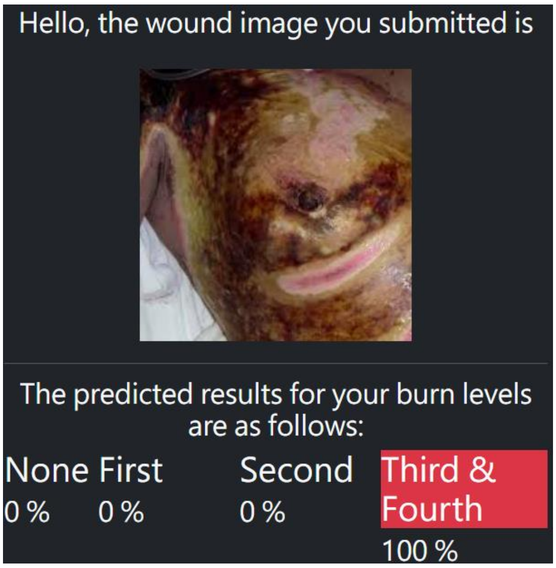


Fig. 3 Schematic diagram of recognition results (third and fourth degree as an example)

3.2 Automated Grading Suggestions

After the system produces the grading result for the image, it will automatically provide current treatment suggestions according to the recognition result and wound type reported by the user. The urgency levels of the suggestions are indicated in different colors, allowing the public to easily distinguish the current severity of the wound. The treatment suggestions are shown in Fig. 4.

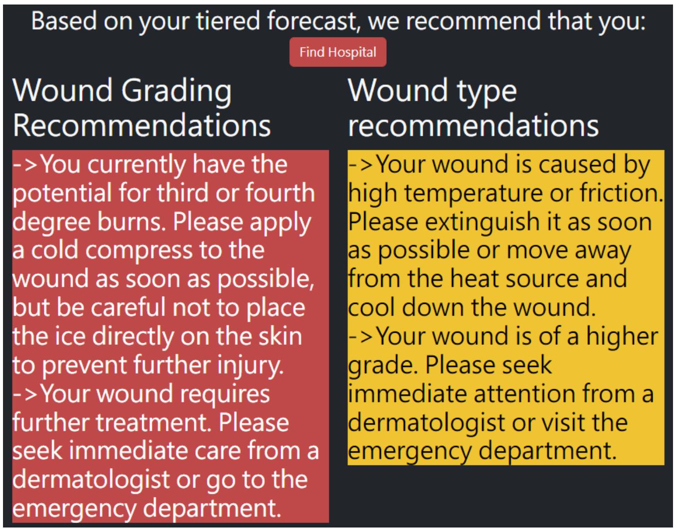


Fig. 4 Schematic diagram of treatment suggestions (third and fourth degree as an example)

3.3 Web Server

After confirming that the model has a certain recognition capability for burns of each grade, we exported the final model and deployed it on a web-based platform. At the same time, in order to provide convenience for medical treatment, the platform also integrates functions such as medical institution lookup and navigation. Users only need to provide the necessary information, and the platform will act as an intermediary to automatically process the data further and return the results to the user. The system architecture of this platform is shown in Fig. 5.

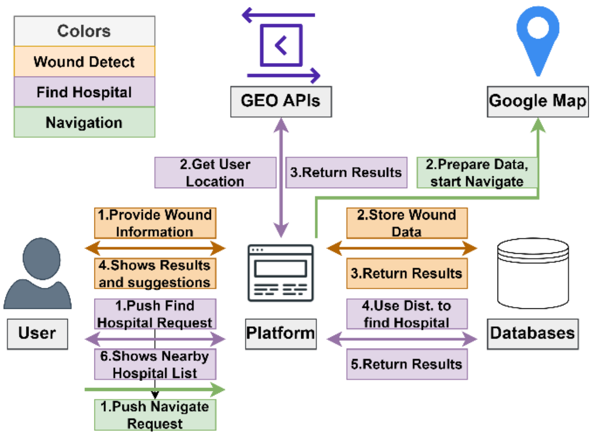


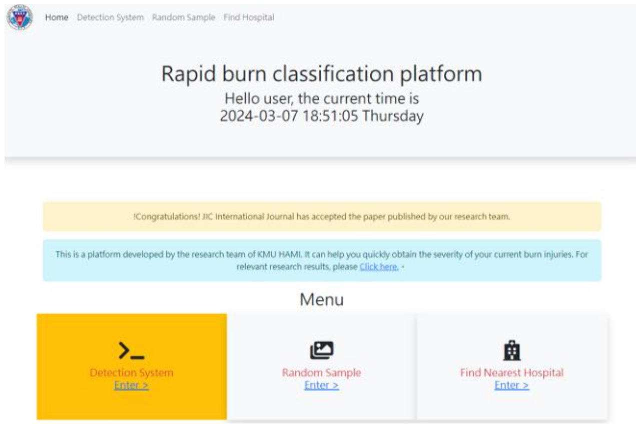
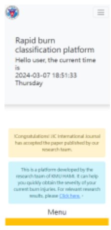
Fig. 5 System architecture diagram

3.3.1 Recognition system interface design

In response to the current trend of users using mobile devices for browsing [9], this platform will use the Bootstrap open-source web framework to allow the layout to automatically adjust according to device size [10]. Table 3 compares the interface layouts on desktop and mobile devices:

Table 3 Layout comparison on different endpoints

Endpoint	Actual screenshot

Desktop	
Mobile	

3.3.2 Wound recognition functionality

In order to make it easy and intuitive for users to use the recognition function, a QR code with interactive data is generated during the image upload stage. Users can scan and upload photos from their mobile phones and operate with their phones in subsequent processes. Fig. 6 shows the full front-end and back-end process flow of the burn recognition function in the system. The brief front-end process flow for users is as follows:

I. Enter the home page (burn.bacons.cc).
II. Click on the recognition system.
III. Confirm wound type.
IV. Upload burn photos from computer or mobile phone.
V. Crop wound photo to appropriate size.
VI. View final grading result.

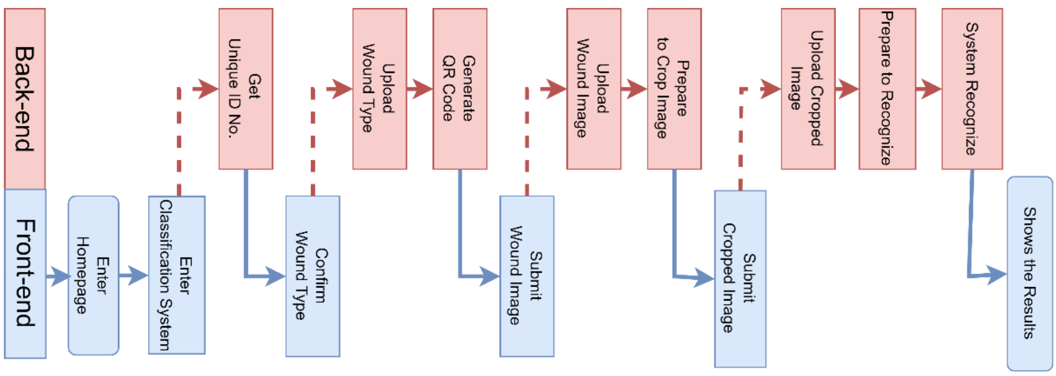


Fig. 6 System front-end and back-end process flowchart of recognition function

3.3.3 Medical institution lookup and navigation functions

When users feel the need for medical treatment, they can use this feature to quickly search for nearby medical institutions with dermatology capabilities. Fig. 7 shows the system's full front-end and back-end process flows for the medical institution lookup function (left half of Fig. 7) and navigation function (right half of Fig. 7). The brief front-end process flow for users is as follows:

I. Enter the lookup function.
II. Provide location permission.
III. View list of surrounding medical institutions.
IV. Select desired medical institution.
V. Redirect out of platform to navigation service.

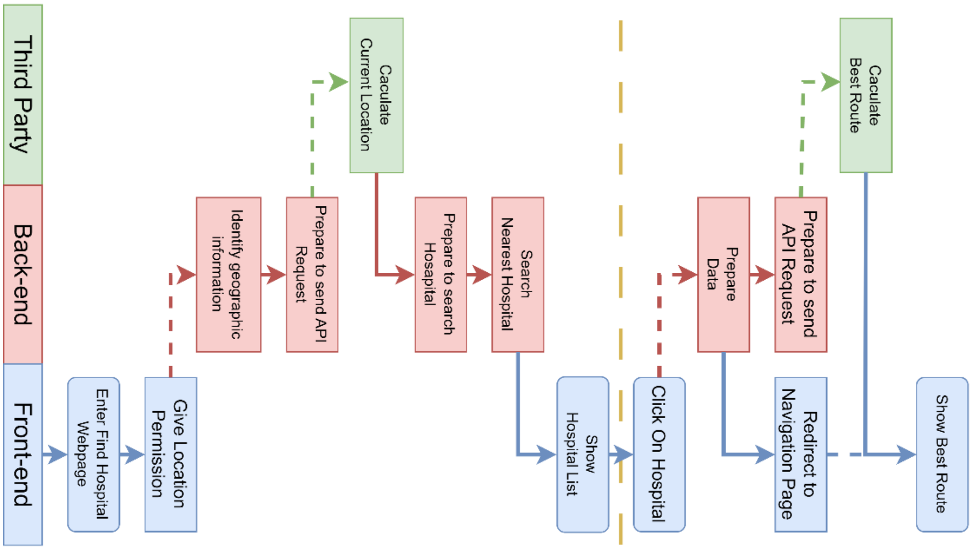


Fig. 7 System front-end and back-end process flowchart of medical institution lookup and navigation functions

3.3.4 Public site information

The recognition platform is now open for public use. The public site information and related platform operation instructions are shown in the table below:

Table 4 Research site and related information

Information	**Link**
Official Platform Website	burn.bacons.cc
Platform Operation Manual	burn.bacons.cc/doc/GL-USGD.pdf
Platform Operation Demo	burn.bacons.cc/demo
Platform Recognition Result Page	burn.bacons.cc/result.php?uid=393

4. Empirical research

4.1 Model Wound Recognition

The recall rate of this model ranged from 81% to 91%. Recall rate refers to the proportion of samples in the test set of this classification that are correctly identified as belonging to that classification. Accordingly, it is known that for some classifications, the model still has omissions to a certain extent that need further improvement. However, the precision results of all classifications are above 81%, indicating that while the model can quickly classify wounds, only a small number of wound images are misclassified by the model due to the combination of multiple wound characteristics of different classifications, or due to excessively low image resolution. In terms of the comprehensive F1-Score, the model performs best in recognizing non-burn images, and still has good predictive effectiveness for images with burns.

4.2 Web Recognition Platform

After the web recognition platform was opened for public use, user feedback on the experience of using this platform has been collected. From mid-July to September 5, 2022, a total of 182 visits were recorded, with 35 valid questionnaires completed. Statistical results show that users' feelings about the system process (Q2-4) and layout (Q2-3) are satisfactory or above, and nearly 77% of users are satisfied or above with the overall system performance (Q2-6). Detailed statistical results for each question are shown in Table 5.

Table 5 Feedback statistics

Question	1	2	3	4	5
Q2-1	0	3	4	23	5
Q2-2	0	0	4	28	3
Q2-3	0	0	13	19	3
Q2-4	0	0	6	24	5
Q2-5	0	4	2	13	16
Q2-6	1	2	5	23	4

Note. 1 represented the worst rating and 5 represented the best rating.

4.2.1 Negative experience

Summarizing the feedback from users who were dissatisfied with the usage experience (i.e. gave any satisfaction rating of 2 or below), the negative experiences can be divided into:

1. Inaccurate grading results

The main cause of dissatisfaction for these users is that the recognition result is very different from the user's expected result, causing the user to feel confused about the recognition result. By tracing back the forms, we found that the images submitted by such users were mostly between the third degree burns, but the platform incorrectly gave a first degree classification result. This allows even users without medical expertise to clearly perceive that the recognition result severely deviates from the medical definition of burns, thus reducing trust and feeling dissatisfied with the usage experience. Follow-up improvements to the recognition model should be continued to increase the accuracy of each classification.

2. Too much text leads to difficulty in reading

Such feedback mostly occurred in users aged 44 and above. The main cause is that for presenting results, whether it is the recognition steps or final results, the platform mainly uses textual descriptions, supplemented by color prompts. This may cause difficulties for older users to read due to too much text, and the platform does not provide other assisted reading functions such as magnifying glass or voice reading. Consideration should be given to how to increase friendliness for the elderly, such as appropriately arranging voice assistance and graphical illustrations.

4.3 Limitations

4.3.1 Wound occlusion

When preparing the training dataset, images that could interfere with the model's feature extraction were excluded in advance. However, in actual situations, a few wounds may still have tattoos, birthmarks and other skin surfaces similar to burns, or the wounds may be occluded by clothing, towels and other fabrics. Whether the model can accurately classify these images remains to be confirmed.

4.3.2 Wound characteristics

Due to the characteristics of burn wounds, the skin appearance may not accurately show the degree of burns within 48 hours, and the wounds might have multiple degrees at the same time. As a result, this platform can only provide the public with a general classification. For actual medical applications, it still needs to be combined with other deep learning techniques [11] to jointly identify the wounds.

5. Conclusion

In this study, we manually reviewed the image dataset, graded the images, and supplemented missing feature images for each classification. The recognition model was trained using the GTM platform, and subsequent effectiveness verification confirmed the model's robust recognition capabilities. When users upload photos, the system automatically generates burn injury grading results and provides medical advice. To enhance user accessibility to the model, we exported the final model and deployed it on a web-based system, incorporating various features and optimizing for mobile use to align with current user habits. Ultimately, nearly 80% of users expressed satisfaction or above with the overall system performance.

Acknowledgments

This research was supported by the National Science and Technology Council, Republic of China under Grant No. 111-2222-E-035-009.

References

[1] Gerstl, J.V.E., Ehsan, A.N., Lassarén, P., Yearley, A., Raykar, N.P., Anderson, G.A., Smith, T.R., Sabapathy, R. and Ranganathan, K. THE GLOBAL MACROECONOMIC BURDEN OF BURN INJURIES. Plastic and reconstructive surgery, 2024, 153(3), 743-752.

[2] Barclay, T.S., Marc, G.J. and Kathryn A.C. Epidemiology, risk factors, and prevention of burn injuries." UpToDate, 2022. Available online: https://medilib.ir/uptodate/show/822.

[3] Adeofe, L. Artificial intelligence and subjective experience. Proceedings of Southcon '95, Fort Lauderdale, FL, USA, 1995, 403-408.

[4] Sofia Z., Maria B.G.Z., Xavier S., Alejandro G., Paul J.K. and Adel E. Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods. Artificial Intelligence in Medicine, 2020, 102. DOI:10.1016/j.artmed.2019.101742.

[5] SHUBHAM, Skin Burn Dataset. Kaggle. 2021.

[6] Mohammad, H. and Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process, 2015, 5(2), 1-11.

[7] Kranthi kumar, How to Measure the performance of a Model? Medium Analytics Vidhya. May 2021.

[8] Zachary Warnes. How to Select a Data Splitting Method. Medium Towards Data Science. June 2021.

[9] Peng, W., Dickson, K.W.C., Kevin, K.W.H. and Patrick, Lo. Why read it on your mobile device? Change in reading habit of electronic magazines for university students. IRE Journals, 2016, 42(6), 664-449.

[10] Suraj S.G. A Review Paper on Bootstrap Framework. International Society for Burn Injuries, in press.

[11] Che W.C., Chun Y.H., Feipei L., Mesakh C., Shih C.H., Dun H.C. and Yo, S.C. Application of multiple deep learning models for automatic burn wound assessment. International Society for Burn Injuries, 2023, 49(5), 1039-1051. DOI:10.1016/j.burns.2022.07.006.