Machine Learning with Applications 9 (2022) 100371

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

A deep convolutional neural network-based approach for detecting burn severity from skin burn images

Sayma Alam Suha, Tahsina Farah Sanam *

Institute of Appropriate Technology, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

ARTICLE INFO

Keywords:
Burn severity
Deep convolutional neural network
Transfer learning
Feature extraction
Image classification

ABSTRACT

Burns being one of the leading causes of clinically significant morbidity can lead to a dramatic physiological reaction with prolonged repercussions, metabolic disturbance, severe scarring, catastrophic organ failure, and death if not properly treated. Appropriate burn treatment management is associated with the severity of burn wounds which can be extremely challenging to anticipate at an early stage due to various factors using traditional clinical methods. Therefore, this study proposed a Deep Convolutional Neural Network (DCNN) based approach for detecting the severity of burn injury utilizing real-time images of skin burns. The DCNN architecture leverage the utilization of transfer learning with fine tuning employing three types of pretrained models on top of multiple convolutional layers with hyperparameter tuning for feature extraction from the images and then a fully connected feed forward neural network to classify the images into three categories according to their burn severity : first, second and third degree burns. In order to validate the efficacy of the suggested strategy, the study also applies a traditional solution to mitigate this multi-class categorization problem, incorporating rigorous digital image processing steps with several conventional machine learning classifiers and then conducts a comparative performance assessment. The study's findings demonstrate that using pretrained models, the recommended DCNN model has gained significantly greater accuracy, with the highest accuracy being obtained using the VGG16 pretrained model for transfer learning with an accuracy of 95.63% . Thus, through the use of intelligent technologies, the proposed DCNN-based technique can aid healthcare practitioners in evaluating the burn damage condition and providing appropriate treatments in the shortest feasible time, remarkably reducing the unfavorable consequences of burns.

1. Introduction

Burns tend to be one of the most prevalent injuries in the world, with consequences that can be deadly or cause a victim to suffer extremely if not treated appropriately. Catastrophic burn injuries are extremely distressing and physically devastating traumas that impact approximately every major organs (Jeschke et al., 2011). According to a report of World Health Organization (WHO), burn injuries cause 180,000 fatalities on average per year, while approximately 11 million people were severely burnt and required medical treatment in the year 2004 (Wang et al., 2018). Radiation, electricity, heat, excessive cold, chemical elements etc. can cause burn injuries, where treatments must be ensured carefully according to its severity (Herndon, Zhang, & Lineaweaver, 2022). With early and adequate treatment, the survival rates of burn victims can be considerably enhanced. Early burn wound excision, skin grafting, skin substitutes are typical treatment techniques that can enhance the prognosis of severe burn patients by lowering fatality rates and minimizing hospital stay days; whereas without correct treatment at the right time poor wound healing, infection, discomfort, hypertrophic scarring, organ failure, and even death can ensue (Stoica, Chircov, & Grumezescu, 2020).

The severity of a burn can be determined depending on the layers of tissues damaged in the human body, with vascular, epidermal, dermal, and muscles among the different tissues typically vulnerable to burn wound (Laggner et al., 2022). The burn injury is usually classified into one of three categories by the healthcare professional: superficial (first degree burns), superficial-partial or deep-partial burns (second degree burns), and full thickness burns (third degree burns), with each category having different healing times and characteristics (Crouzet et al., 2015). Burn wounds are dynamic and can develop as well as convert to deeper wounds, making an accurate estimate of their depth and severity extremely challenging at an early stage (Rice & Orgill, 2021). Sufficient functional and structural investigations are necessary for precise burn intensity diagnosis. To measure the burn severity, modern techniques such as laser Doppler imaging or medical evaluation under the supervision of experienced healthcare practitioners are necessary in traditional clinical practice, but these procedures are constrained by factors such as availability of the devices, distance, time, expense etc (Rangaraju et al.,

* Corresponding author.
E-mail addresses: suha.mist@gmail.com (S.A. Suha), tahsina@iat.buet.ac.bd (T.F. Sanam).

https://doi.org/10.1016/j.mlwa.2022.100371
Received 18 April 2022; Received in revised form 8 June 2022; Accepted 16 June 2022
Available online 30 June 2022

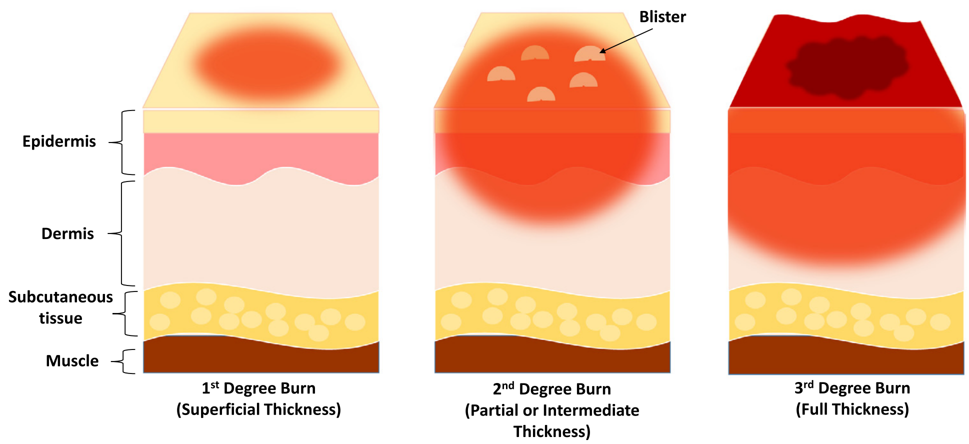


Fig. 1. Burn depth classification.

2019). Because of these obstacles, the burn victims' treatment process may be delayed, resulting in a severe health deterioration. Furthermore, standard manual methods for estimating burn severity, such as visual inspection and physical assessment, not only cause delays, but have also been proven to provide estimations that are only 50%–70% correct during the early days after a burn (Chauhan & Goyal, 2020). Early detection of burn depth severity becomes more challenging in remote areas of least developed and developing countries, where healthcare resources and facilities are scarce.

Therefore, the purpose of this study is to propose an autonomous classifier that can detect burn severity from real-time photographs of the skin burn area and categorize it as first, second, or third degree burns. To attain this objective, a Deep Convolutional Neural Network based machine learning classification model has been designed, trained, and tested using 1530 images of skin burns which intends to use burn photos for detecting and categorizing patients' burn severity. The significant contributions to achieve the goal of this study are listed below.

- A Deep Convolutional Neural Network (DCNN) based approach has been proposed where a Convolutional Neural Network incorporating different state-of-the-art techniques like transfer learning with pre-trained models and fine tuning on top of multiple convolutional layers with hyperparameter tuning have been employed for feature extraction from victims' real-time burn photos; and then a fully connected Artificial Neural Network has been used to classify them according to their severity into first, second and third degree burns.
- The classification of the burn images according to their severity have also been conducted through another approach employing traditional machine learning technique where the feature extraction have been conducted meticulously through several image processing stages and then the classification has been conducted using six types of conventional machine learning classifiers.
- To validate the efficacy and potency of the proposed DCNN technique, a comparative analysis have been conducted between the traditional approach and suggested method. Also, three kinds of pre-trained models have been employed and tested in the CNN architecture with an aim to explore the best performing model in this scenario.

The remaining paper is organized as follows: Section 2 explores the background study, Section 3 demonstrates the research methodology; Section 4 discusses the result analysis and findings; and Section 5 concludes with a discussion and conclusion that highlights the important findings and future research plan.

2. Background study

Burn damage is a typical occurrence in which a deep and extensive burn can result in catastrophic consequences such as sepsis from bacterial infection, shock from hypovolemia, massive fluid loss, organ failure, and so on if not treated early (Shpichka et al., 2019). Burn injury can be classified based on its severity, depth of burn and size. Burns that just damage the top layer of the skin called epidermis are classified as superficial or first-degree burns in which the skin turns red and the pain is short-lived; partial or intermediate thickness known as second-degree burns are painful, drier, creates blisters, require dressing with wound care, and may scar, but they do not typically necessitate surgery; and finally a full-thickness or third-degree burns are dry that go through the entire dermis and is usually not painful due to nerve loss, but it does requires fluid resuscitation, protection from infection, and unless the burn is extremely minor surgical care is essential (Jeschke et al., 2020; Noorbakhsh, Bonar, Polinski, & Amin, 2021). Fig. 1 shows the illustration of categorization for burn depth degrees according to its severity (Jeschke et al., 2020). Along with clinical examination, Laser Doppler based techniques such as Laser speckle imaging (LSI) or Laser Speckle Contrast Analysis (LASCA); thermal imaging; Spatial Frequency Domain Imaging (SFDI) etc. are among prominent techniques in medical field for correctly assessing perfusion in burns and burn depth detection (Ponticorvo et al., 2019). Unfortunately, these procedures need the supervision of qualified specialists, who may not be accessible at the time of the burn injury and thus the burn wound progression may occur rapidly.

To solve this issue, several researches worldwide have applied various computational techniques to automatically classify the burn images and predict the severity of burn damage from the captured injury images in real-time. In case of image classification based tasks, machine learning approaches are one of the most extensively utilized and promising techniques, which generally analyze and retrieve critical information from enormous quantities of heterogeneous data in order to detect and classify anomalies autonomously (Tchito Tchapga, Mih, Tchagna Kouanou, Fozin Fonzin, Kuetche Fogang, Mezatio, & Tchiotsop, 2021). Therefore, employing various machine learning techniques for burn severity assessment is gaining traction nowadays. For example, the study referenced in Şevik, Karakullukçu, Berber, Akbaş, and Türkyılmaz (2019), used 105 burnt photos to develop an automatic segmentation-based classification method to categorize burn images into healthy skin, burned skin, and backgrounds for which they employed four types of clustering approaches for image segmentation and then applied several traditional machine learning classification techniques with an aim to explore the best performing classifier. In the paper (Kuan, Chua, Safawi, Wang, & Tiong, 2017), an image mining strategy was used to categorize different burn levels of captured burn images into three groups utilizing a comparative evaluation of 20 types of machine learning classification algorithms using both test dataset and 10 fold cross validation approach. In the Study referenced (Yadav, Sharma, Singh, & Goyal, 2019), the authors utilized 74 burn images to develop a feature extraction model with several digital image processing steps and then classified the images into two classes using Support Vector Machine classifier. Another related work had been explored in study Rowland et al. (2019), where a spatial frequency-domain imaging (SFDI) approach was combined with a support vector machine (SVM) machine learning classifier to develop a model that can predict severity of progressive burns in a pig model and to estimate burn severity by measuring the absorbance and scattering characteristics of burn tissue. The work Hai, Triet, Thai, and Thuy (2017) had proposed a method for categorizing burn photos into the second, third, and fourth degrees of severity, in which they used a combination of image processing techniques concentrating on color feature extraction from the images and then SVM classifiers to categorize the images. Another work in Lee et al. (2020) suggested a real-time technique for classification of burn depth employing moderate sample sizes based on ultrasound imaging, in which the textural feature set is constructed using a grey-level co-occurrence matrix (GLCM) derived from the ultrasound pictures of the burn tissue; and then utilizing a nonlinear support vector machine and kernel Fisher discriminant analysis, classification is accomplished in porcine skin tissue under four different burn scenarios.

Recently a few studies also applied deep learning technique in this research area for classification and automatic severity detection of burns. For example, the study referenced in Karthik, Nath, and Veena (2021) proposed a predictive model based on deep neural network, recurrent neural network(RNN) and CNN to determine degree 1, degree 2 and degree 3 of burn images depending on the severity of the burn over a dataset of 104 images. In another study Pabitha and Vanathi (2021), the authors presented a DenseMask Regional convolutional neural network technique, which combined a Mask-region based convolution neural network with dense pose estimation for segmenting the Region of Interest of a skin burn areas from images based on the severity of the burn damage. Another work proposed in the paper Abubakar, Ugail, Smith, Bukar, and Elmahmudi (2020), applied deep neural network with transfer learning using two pre-trained models ResNet50 and VGG16 for the feature extraction from images and then applied SVM classification approach to classify the images into four categories which are healthy skin, first degree, second degree and third-degree burns over 2080 RGB input images. The authors developed a deep learning-based system in work Liu, Yue, Cheng, Li, and Fu (2021), that included precise burn area segmentation and burn depth labeling, as well as proposed a framework for enhanced burn area segmentation and automated burn depth diagnosis based on deep learning methods. Also, the study referenced in Tran, Le, and Nguyen (2016) suggested an approach for skin burn depth identification in which the pictures are pre-processed using Local Binary Pattern (LBP) operations based on recommendations of a burn specialist, and then an adaptive CNN architecture is used to categorize burn images into four degrees based on their severity.

However, in the earlier studies rarely any researchers have focused on the efficacy of employing CNN architecture with deep neural network over the traditional method of image classification with feature extraction via image processing techniques and conventional machine learning classifiers. Also, the utilization of transfer learning method through various pre-trained models with fine tuning have been explored little. Therefore, this study proposes a CNN architecture that integrates several state-of-art techniques such as transfer learning with fine tuning to effectively classify the burn images according to their severity; and also it conducts the classification analysis through traditional approaches to have a comparative study with the proposed method.

3. Methodology

In this research, a Deep Convolutional Neural Network (CNN) based machine learning classification technique has been proposed, trained and tested that aims to differentiate between the skin burn depth degree of patients utilizing the burn images. Here, the burnt photographs of the victims, captured using a digital camera or a mobile phone camera, have been used as the study's input, with the system predicting the degree of the skin burn. Along with the proposed CNN technique the traditional image classification approach incorporating digital image processing and conventional machine learning classifiers have also been explored in this study for skin burn depth detection in order to assess a comparative performance and efficacy analysis of the suggested technique. The framework of the research methodology followed in this study has been illustrated in Fig. 2. The stages involved are briefly discussed in the following sections.

3.1. Data acquisition

The images utilized in this study are digital color images of skin burns captured with camera. There are total 1530 images of skin burns from which 1440 images have been obtained from a publicly available repository of Kaggle (2022) and the rest of the images have been collected from the burn units of three hospitals of Bangladesh maintaining their patient privacy concerns. The photographs which have been taken directly from the hospitals were captured using three distinct mobile phones containing digital camera. All of the pictures from the open source repository, as well as those taken directly from the hospitals, are color pictures captured with an emphasis on the burn regions of the victims. Additionally, the images have optimal lighting conditions and good picture quality, allowing the burnt and unburned parts to be distinguished from the background environment. According to their severity, the images are labeled as first degree, second degree, or third degree burns under the supervision of four healthcare practitioner with two burn specialist doctors, where first degree burn photographs depicting less damage and third degree burn being severely damaged. After categorizing them, it has been observed that the dataset contains 670 images of first degree, 347 images of second degree and 513 images of third degree burns.

Here, as the digital input images are all RGB colored images, the number of channels for each picture is 3. The pictures' formats and sizes vary since they were retrieved from various sources, rendering them inappropriate for using in predictive analytics. To resolve this issue all the images have been resized to the shape $(224 \times 224 \times 3)$ that contains information about the row, column, and channel of the image and then stored them in a 3D multichannel array. The labeling information for each image has also been stored in another one dimensional array which will be used as the target attribute for the machine learning phase. The examples of one image from each category of input images are shown in Fig. 3.

3.2. Data augmentation

Data augmentation, which comprises generating slightly modified duplicates of images from the original training samples, is an effective strategy for reducing data scarcity as well as improving performance and minimizing forecast error (Shorten & Khoshgoftaar, 2019). In this study, three types of basic image manipulations have been applied to each images to formulate four additional training samples with image augmentation process which includes geometric transformations (random flipping, rotating and shifting); contrast enhancement and sharpening with noise removal. As a consequence, the dataset without augmentation has 1530 images, and a new augmented dataset containing 6120 photographs of burnt areas has been generated applying augmentation techniques.

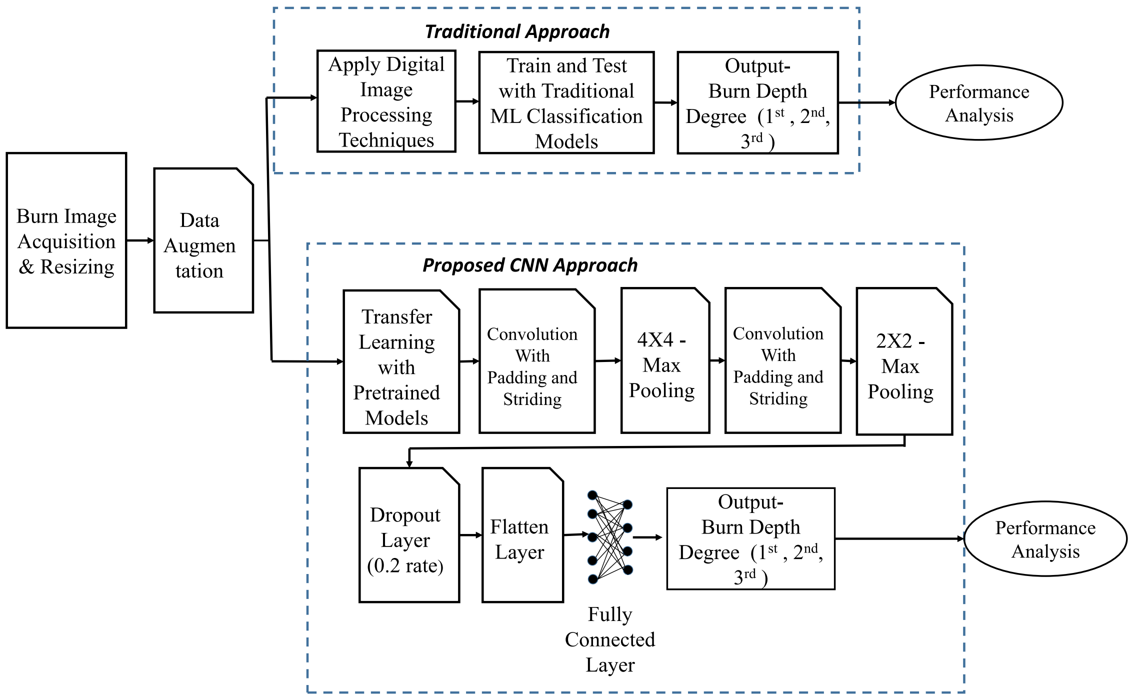


Fig. 2. Framework of research methodology.

Fig. 3. Example of first, second and third degree burnt input images.

3.3. Applying machine learning techniques

In this research, two types of machine learning techniques have been employed to classify the burnt images of patients, where the first one is with several conventional machine learning classifiers and the second one is with the proposed CNN architecture incorporating several state-of-art techniques like transfer learning, fine tuning etc. The Google Colaboratory has been utilized here as the platform for execution, with the scikit-learn and TensorFlow packages from Python. The two approaches are described hereafter.

3.3.1. Traditional approach

Training with a large number of high-resolution images necessitates the processing of large volumes of data, which is troublesome for most traditional machine learning models (Lézoray, Charrier, Cardot, & Lefèvre, 2008). For this reason, it is required to perform the appropriate digital image processing steps to extract and highlight the key information from the images before utilizing them in the conventional machine learning classifiers. Thus, in this approach the images are initially preprocessed with digital image processing steps; then splitted into train and test data; and finally employed in the traditional machine learning classifiers.

- **Digital Image Processing Steps:** After the image acquisition and resizing, the first step of digital image processing that has been applied here is the "Image Enhancement". This is achieved here via the histogram equalization approach, which enhances the visual perception of information in photographs by giving a more evenly dispersed histogram, resulting in a sharper and clearer image for viewers. As histogram equalization operation cannot be done in colored RGB image space, thus the images are converted to Hue Saturation Value (HSV) format, apply histogram equalization in that HSV image and then again converted back to RGB format which will resulted into an enhanced image. The next stage is "Noise Removal" for which as an image restoration phase of noise reduction, a median filter approach is used; which is a commonly used nonlinear filter with superior edge maintaining capabilities and the potential to minimize impulsive noise (Zhu & Huang, 2012). Following that, the required amount of "Morphological Erosion" and "Morphological Dilation" operations have been performed where erosion reduces and dilation enhances the pixel numbers on the edges of objects in the images. The final image processing step applied here is "Image Segmentation" to divide the images into different significant areas. For this, here k-means unsupervised clustering algorithm has been applied that segments the interest area of burn from the background and thus separates the affected area from the healthy skin (Kumari, Prasad, & Mounika, 2019).
- **Split into Train and Test Data:** After the specified image processing stages are completed, the set of processed pictures are turned into a single dimension array by using OpenCV python function 'FLATTEN'. Each column in this array represents an attribute that contains the image's pixel values. This array with the pixel values has been used as the training's input features, with the one-dimensional array containing the labeling information for each image functioning as the target attribute. Finally, to use this processed dataset in the machine learning classification models, the dataset is divided into 70% train data and 30% test data.
- **Training with Traditional Machine Learning Classifiers:** This study's predictive analysis is a multiclass classification problem where the images can be classified into three classes. Therefore, six different types of conventional classifiers have been employed here with an aim to classify the images into first, second or third degree burn. The classification models are: Logistic Regression classifier, K-Nearest Neighbor(KNN) classifier, Support Vector Machine (SVM) classifier, Decision Tree classifier, Random Forest classifier and Multi-layer Perceptron classifier.

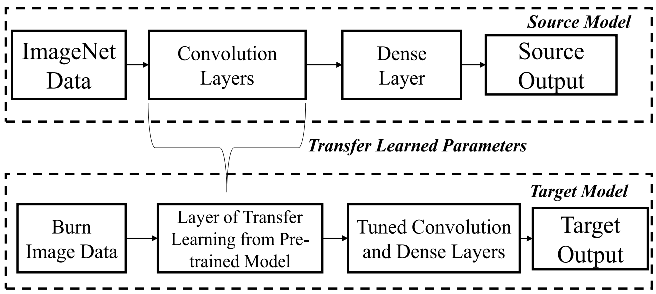


Fig. 4. Basic framework of transfer learning method.

Therefore, in the traditional machine learning approach these above mentioned steps are followed in order to classify the images according to their burn depth severity.

3.3.2. Proposed CNN approach

Convolutional Neural Networks (CNN) are indeed a type of deep neural network with multiple consecutive layers that have shown to perform efficiently in a range of image processing, classification, and segmentation tasks (Kim, Jung, Park, Lee, & Ahn, 2022). The advantage of employing CNN is that, it can successfully tackle picture classification problems with higher accuracy since it matches the data point distribution in the picture throughout the neural network training process and can directly utilize the feature maps from the convolutional layers. As a consequence, substantial features from the images can be extracted autonomously without the need for explicit image processing operations (Dabeer, Khan, & Islam, 2019).

In this research, a CNN architecture with different layers was rigorously built with the goal of classifying burn images. After data acquisition from the repository, the dataset is divided into 70% train and 30% test data in this approach; and then the CNN model is built to use this dataset for training. "Sequential" is the model type that has been employed here to construct the CNN architecture layer by layer with the help of "add()" function for adding each layer. The layers of the CNN architecture are discussed hereafter.

- **Transfer Learning with Pretrained models** The architecture's first layer after data augmentation is based on transfer learning with fine tuning, a powerful machine learning strategy that aims to improve target learners' performance on intended domains by transferring information from the relevant pre-trained models and then making small adjustment to get the prediction (Zhuang et al., 2020). Here, the pre-trained model's fully connected or dense layer from the source task has been removed using "$include\,top = False$" operation and retrained them for the target task using some more convolution and dense layers with hyper-parameter tuning. The basic framework followed for transfer learning has been illustrated in 4.
  In this study, three different types of pre-trained models have been trained, and tested in the transfer learning layer to evaluate which model provides better performance for classification. The models are: VGGNet16 model; MobileNet model; and ResNet50 model (Keras, 2022). The Tensorflow and Keras library contains pre-trained models, where the weights are derived from three-channel pictures. These pre-trained models have been trained using millions of images from the ImageNet dataset to predict over 1000 classes; consequently, utilizing pre-trained models allows a new model to converge quicker and perform better on a smaller dataset by leveraging features learnt on a bigger dataset (Arias-Garzón et al., 2021). The pretrained models are here used without their final classification layer, which will aid in the transformation of images from a new domain job based on its hidden states, allowing for the extraction of features from a new domain work while leveraging information from a source-domain task.
  VGG16: The first pretrained model that has been used in this study is VGG16 containing 16 convolution layers with different weights which follows the architecture of containing 3x3 filters with a stride 1 and the same padding and maxpool layer of a 2x2 filter with a stride 2 as well as contains two fully connected dense layers in the end, followed by a softmax for output. This network is quite vast, with approximately 138 million parameters. This model has provided best accuracy of 90.1% with Imagenet dataset (Tammina, 2019)
  MobileNet: MobileNet is a basic yet efficient convolutional neural network that is commonly used for mobile vision applications has been employed in this study as one of the pretrained models. The MobileNet Architecture is built on the foundation of two types of convolution layers: depthwise separable convolution and pointwise convolution layers. The depthwise layer is the map of a single convolution upon every input channel individually, followed by pointwise convolution layers with a 1x1 filter multiplication operation to merge the depthwise layer's feature map results. This model has provided best accuracy of 89.5% with Imagenet dataset (Harjoseputro, Yuda, Danukusumo, et al., 2020)
  ResNet50: The third type of pretrained model that has been employed here is ResNet50 which had been introduced based on the concept of residual network to solve the problem of vanishing gradient in deep neural network. ResNet50 comprises 50 layers of residual networks, which contains distinct groups of identical layers, with identity blocks connecting two layers of varied sizes in between. The Skip Connections between layers function adds prior layer outputs to the outcomes of stacked layers, allowing for far deeper network training than formerly feasible. This model has provided best accuracy of 94.5% with Imagenet dataset (Mukti & Biswas, 2019)
- **Convolution Layer with padding and striding** CNN's essential and fundamental component layer is the next layer which is called "Convolution layer" that consists of a series of kernels or filters that are mostly smaller in size than the original training picture and whose parameters must be trained throughout time of learning and convolve with the real image (Mostafa & Wu, 2021). Convolutional layers use a collection of convolutional kernels to extract features, which perform convolution operations on the input dataset or feature maps through intermediate layers (Ren et al., 2022).
  In this study after the transfer learning layer with pre-trained model, the 2 dimensional convolution layer (Conv2D) operation has been performed two times where the first layer contained 128 nodes and second one 64 nodes. In both cases, the kernel or filter size for convolution has been considered to be 3×3. Moreover, "padding" and "striding" techniques have been used here in the convolution layer to enhance the accuracy of the output; where padding adds one additional layer to the outer image and striding manages the space between two consecutive kernel locations. The value of padding in this CNN architecture is "same" which indicates padding the input image with zeros evenly to the outer side. The value of stride is 1, which signifies the output size after convolution operation with the filter will be the same as the input size. Then, an activation function is employed in this deep neural network to feed a weighted sum of input signals through and with the result being utilized as an input to the next layer. In this CNN model, "Softmax" activation function has been employed which tends to produce useful outcomes that can be applied to multiclass classification problems and provides an output between 0 to 1 range representing the probability classification outcome (Sharma, 2019). The mathematical representation of the softmax activation function has been shown in Eq. (1).

$$Softmax\ Activation, S(a_i) = \frac{exp(a_i)}{\sum_{j=1}^{n} exp(a_j)} \tag{1}$$

- **Pooling Layer:** The "Pooling Layer" is the following layer in this study's CNN design, which declines the spatial size of the picture while retaining crucial data to progressively reduce the computations and spatial size in the complex and large neural network (Yamashita, Nishio, Do, & Togashi, 2018). Here, a "Max Pooling" technique has been used to obtain the maximum value for each patch of the feature map. The pooling layer has been employed two times after each of the convolution layers with the first max-pool layer having a pooling window size of

  4×4

  and the second max-pool layer having a pooling window size of

  2×2.
- **Dropout Layer:** The "Dropout Layer" is the next layer, which is a regularization method that significantly reduces overfitting and speed up the learning process by changing input data to 0 for ignoring some nodes at a predetermined frequency rate at each step during the training process (Nandini, Kumar, & Chidananda, 2021). To conduct the regularization, the dropout rate in this CNN model has been set to 0.2 in the dropout layer.
- **Flatten Layer:** The "Flatten layer" is the next layer, which converts the multi-dimensional outputs of the preceding layer into a one dimensional array to be used in the next classification layer. The input layer of the classifying neural network is built using this one-dimensional array, with the components of the array being provided to each neuron. Therefore, this layer works as the bridge between the convolution and dense layer. The most significant features from the images are also extracted in this layer as a result of the previous layers.
- **Fully Connected Layer:** The "Fully Connected Layer" also known as "Dense Layer" is the classifier and final layer of this CNN architecture. This layer is at the bottom of the CNN model, and every neuron inside it is connected to every neurons in the previous and forward layers, adopting the standard multiple-layer perceptron and feed forward neural network technique (Alzubaidi et al., 2021). As the classification problem here has three classes, the last layer of this dense network will contain three nodes to provide the classification prediction, one for each possible outcome. Here also "softmax" activation function has been utilized to generate the result from the output neurons possessing the highest possibility.
- **Compiling Model:** Finally, the CNN model has been compiled using three parameters: optimizer, metrics and loss. The optimizer utilized here is the stochastic gradient descent (SGD) optimizer, which is used to regulate the learning rate. Here the learning rate is 0.01. "Accuracy" has been used as the metrics of the model for assessing the performance of the training. And lastly, for evaluating the loss "*categorical_crossentropy*" function is used as the problem here is a multi-class classification problem. The lower value of loss indicates better performance.

To explore the best performing classification model, the CNN architecture formulated in the above mentioned way is trained with the dataset in four different methods and hyperparameter optimization : (i) without employing transfer learning layer; (ii) incorporating "VGG16"; (iii) incorporating "MobileNet"; and (iv) incorporating "ResNet50" pretrained model for transfer learning layer. For each type the training has been conducted in 30 epochs and then the models have been evaluated with test dataset.

3.4. Performance analysis

To evaluate the efficacy of the predictive analysis, the performances of both types of machine learning approaches are assessed using four performance metrics which are the Accuracy, Precision, Sensitivity (recall) and F1 score. The performance indicators are mostly based on a comparison of forecasted and real values from the training dataset, which is separated into four groups: True Positive (TP) that refers to a situation in which both the true and predicted values are positive; True Negative (TN) in which the original value is negative and also the anticipated value is negative.; False Positive (FP) a scenario in which the real value is negative but the anticipated result from the training is positive and finally the actual value is positive, but the forecasted result is negative, leading in a False Negative (FN). The performance metrics can be expressed as following Eqs. (2),(3),(4) and (5) based on these assessments:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \tag{4}$$

$$F1 - score = \frac{2 * (Precision * Sensitivity)}{Precision + Sensitivity} \tag{5}$$

4. Result analysis

4.1. Findings from the traditional machine learning approach

Employing the traditional ML approach, the images have been preprocessed with required digital image processing stages before applying them into ML classification algorithms. One of the examples of a burnt image with each steps of image processing technique has been illustrated in Fig. 5. From the figure it is visible that, after applying the digital image processing steps the affected burnt area is more prominently detected in the final segmented image in comparison to the original image. Similarly, all the images are processed in this way and then employed to the machine learning classifier as train and test image data.

Both types of datasets have been used to train, test, and assess the six types of standard machine learning classifiers: dataset without image augmentation having 1530 images and dataset with image augmentation containing 6120 images. The performances of the six types of machine learning classifiers with different performance metrices employing two types of dataset have been shown in Table 1. From the performance analysis in Table 1, it can be observed that, for all the classifiers the performances enhance when they are trained with augmented dataset. For example, the accuracy of Decision Tree classifier without augmented dataset is 70.6%, which becomes 71.4% after employing dataset with augmented images. This results indicate that, the prediction performances of the machine learning models improves when an increased amount of data are utilized to train the models.

Furthermore, among all the classifiers, Random Forest classification model comparatively outperforms others in terms of performing with highest accuracy being 77.3% without augmented dataset and 80.4% with augmented images whereas the Multi-Layer Perceptron provides the least accuracy 53.0% without augmented data and 56.0% employing augmented image dataset.

4.2. Findings from the proposed CNN approach

The proposed CNN architecture has been slightly modified with hyperparameter tuning in four different ways to explore the best performing model for burn depth prediction where each method has executed over 30 epochs for training. Fig. 6 shows the accuracy and loss per epoch of the training with 30 epochs where Fig. 6(a) shows the accuracy and loss for the CNN model without transfer learning, 6(b) shows CNN model with VGG16, 6(c) shows CNN model with MobileNet and 6(d) shows CNN model with ResNet50. From the figure it is apparent that for all the models the accuracy gradually increases and the loss decreases. Also, the accuracies and losses fluctuate slightly

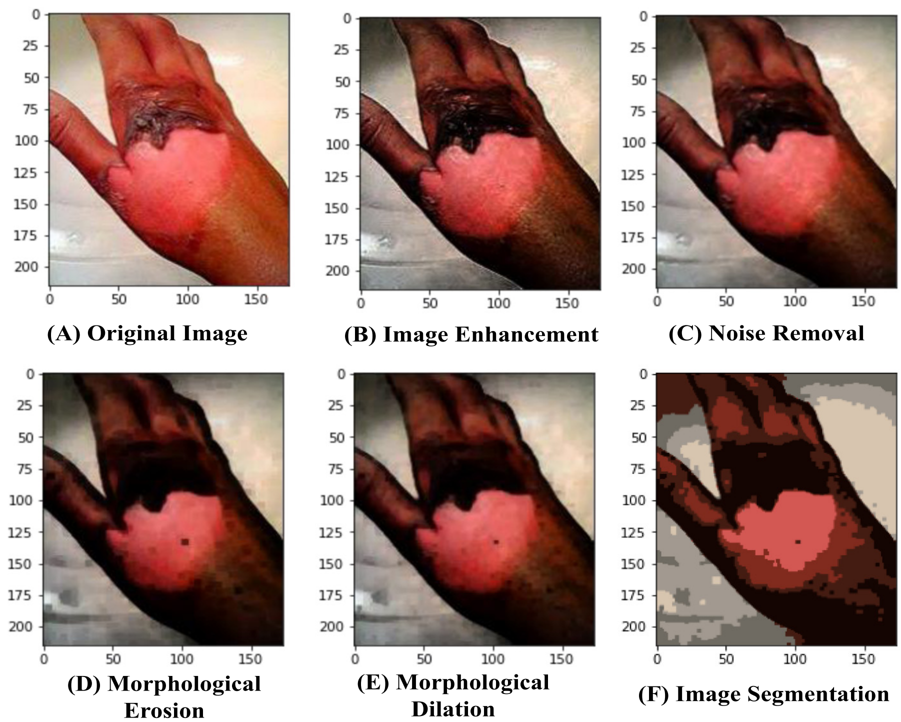


Fig. 5. Findings from Image Processing Stages.

Table 1
Performance analysis results obtained from the test data employing traditional machine learning approach with and without augmented dataset.

Machine Learning Models	Without Augmented Dataset				With Augmented Dataset			
	Acc.	Prec.	Rec.	F1-sc.	Acc.	Prec.	Rec.	F1-sc.
Random Forest Classifier	0.773	0.784	0.734	0.744	0.804	0.81	0..807	0.809
Support Vector Machine	0.767	0.762	0.740	0.747	0.78	0.785	0.781	0.778
Decision Tree	0.706	0.692	0.692	0.692	0.714	0.715	0.725	0.712
K-Nearest Neighbor	0.661	0.796	0.571	0.517	0.696	0.698	0.698	0.684
Logistic Regression	0.642	0.619	0.601	0.601	0.653	0.654	0.655	0.623
Multi-Layer Perceptron	0.530	0.595	0.580	0.524	0.56	0.565	0.564	0.543

Table 2
Performance analysis results obtained from the test data with proposed CNN machine learning approach.

CNN Architecture	Accuracy	Precision	Recall	F1-score
CNN without Transfer Learning	0.7715	0.762	0.749	0.780
CNN with Transfer Learning- (VGG16)	0.9580	0.96	0.95	0.95
CNN with Transfer Learning- (MobileNet)	0.8992	0.94	0.89	0.90
CNN with Transfer Learning- (ResNet50)	0.9325	0.93	0.91	0.92

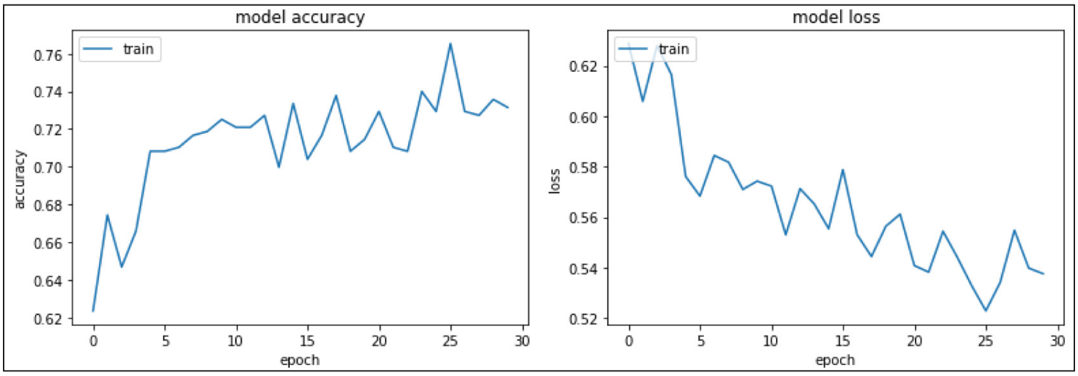
after 25 epochs for all the models, indicating that the models have converged. Here, among all the models Fig. 6(b) illustrating accuracy and loss of CNN with transfer learning employing VGG16 pretrained model shows the best performance where the accuracy and loss converge comparatively earlier being stable at almost 99% accuracy and 0% loss. On the other hand, Fig. 6(a) representing the accuracy and loss of CNN model without transfer learning shows the worst performance with frequent fluctuations of accuracy and losses as well as the accuracy being

Table 2 shows the performances of the proposed CNN technique on the test dataset with four methods; which consists of without including transfer learning from pretrained models as well as including three different pre-trained models as transfer learning process. From the table, it is apparent that all of the CNN approaches with the proposed technique have achieved substantially higher accuracy, with the transfer learning model using the "VGG16" pre-trained model outperforming the other models with 95.63% accuracy. Moreover, it is also noticeable that, the performance of the classification model enhances significantly while employing transfer learning technique as the CNN model without transfer learning has acquired only 72.01% accuracy.

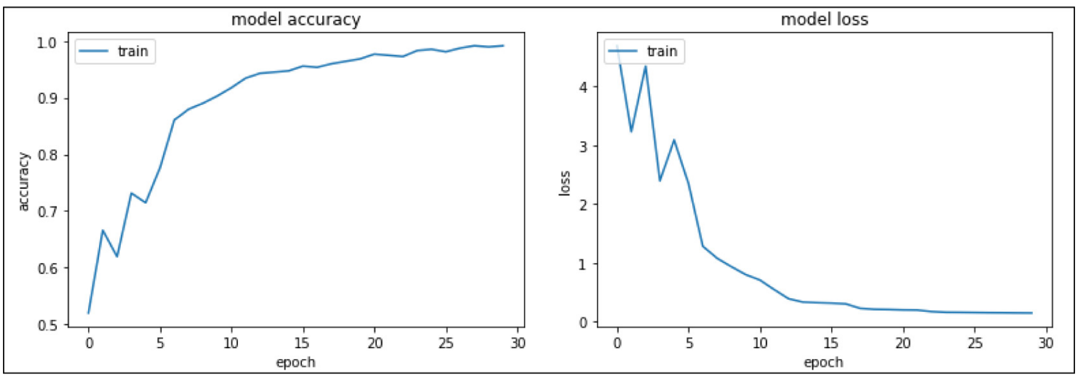
The comparative accuracy analysis of employing traditional and proposed CNN technique with different machine learning models have been illustrated in Fig. 7. From the comparative performance analysis it is evident that, the accuracy of the models with traditional approaches is much lower in comparison to the CNN techniques. Furthermore, the traditional technique entails several tedious image processing phases that need meticulous adjustments whereas the CNN technique can provide excellent performance without any explicit image processing steps as the neural network itself extract the significant features from the images at the time of training. Therefore, the suggested method, which integrates transfer learning with pre-trained VGG16 model, can be used effectively in clinical practice to estimate the depth of burn from captured photos of skin burn injury.

4.3. Limitations and future perspectives

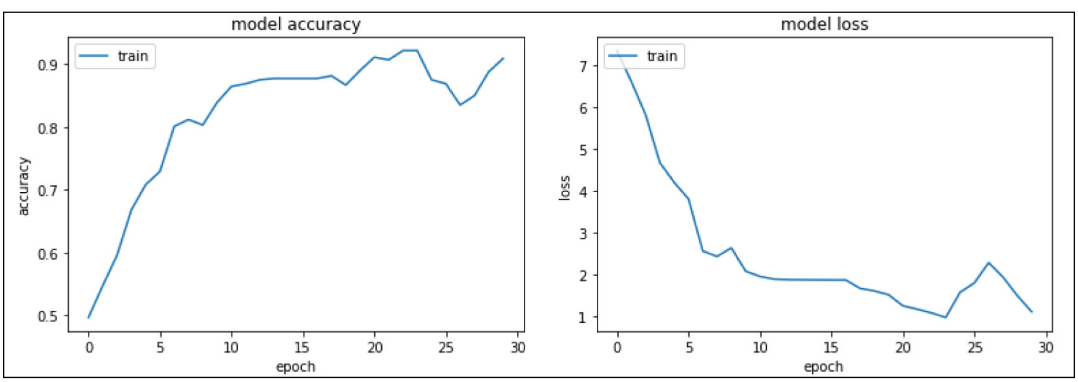
One of the research's limitations is that, although the study has employed photographs from a variety of archives, almost all of them have precise views of the victims' burn regions. It is recommended that clear and sharp photographs should be captured focusing on the burn region of the patient that is distinguishable from the background, with decent lighting, reduced shadows, and moderate or good color picture



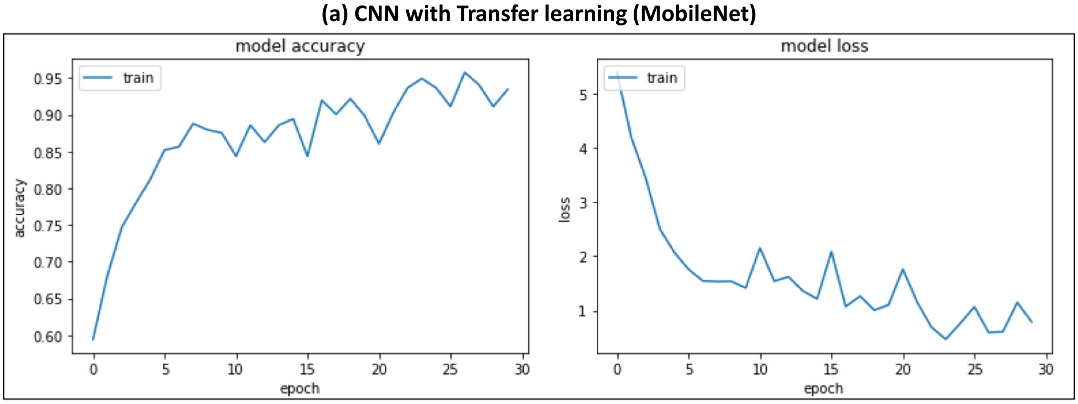
(a) CNN without Transfer learning



(a) CNN with Transfer learning (VGG16)



(a) CNN with Transfer learning (MobileNet)



(a) CNN with Transfer learning (ResNet50)

Fig. 6. Accuracy and loss obtained per epoch from (a) CNN without transfer learning (b) CNN with VGG16 pretrained model for transfer learning (c) CNN with MobileNet pretrained model for transfer learning (d) CNN with ResNet50 pretrained model for transfer learning.

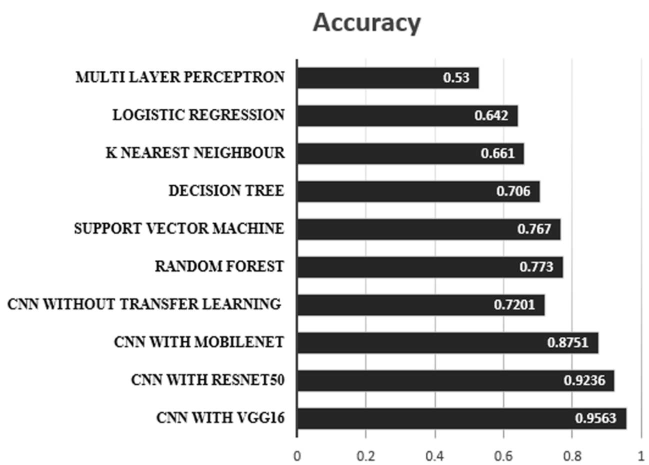


Fig. 7. Comparative accuracy analysis.

quality for this study. However, in a real-time setting, there may be photographs acquired by users with blurry views of burns and lower picture quality that may become difficult to classify in terms of burn severity level. Thus, in future, the research team intends to perform some practical post predictive analysis with different sample of data, focusing on this flaw in the deployment of a real-time predictive model in clinical sector based on the proposed CNN classifier to ensure that all instances can be reliably anticipated to predict the burn severity of patients.

Moreover, this study has utilized only three types of pretrained models for transfer learning in the CNN architecture. Also, for intelligent clinical applications, an explainable AI plays a significant role providing an explanation with proper justification of predictions made by the AI systems, which has not been incorporated in this study. Therefore, in future, the researchers plan to employ more varieties of potential pretrained models with enhanced convolution layers as well as aggregate the techniques of XAI (explainable AI) with the proposed methodology in order to explore a better performing classifier with clinical validation.

5. Discussion and conclusion

This research has proposed a DCNN based model with transfer learning and fine tuning for effective identification of skin burn degrees according to their severity from real-time burn images of victims. The architecture employs transfer learning from pretrained models and then multiple convolutional layers with hyperparameter tuning for feature extraction from the images which are then utilized in a fully connected feed forward neural network to classify the images into three categories. To validate and evaluate the efficacy of the proposed technique, the traditional approach incorporating digital image processing and conventional machine learning classifiers has also been applied to solve this multi-class classification problem. Accuracy, precision, recall and F1-scores are the metrices that have been utilized to conduct the performance analysis. The findings from comparative analysis indicates that the proposed technique shows much higher accuracy in comparison to the traditional approach and also incorporating transfer learning outperforms other techniques in terms of accuracy.

Therefore, the suggested DCNN based computational model can assist medical practitioners and healthcare providers in evaluating the injury condition and suggesting suitable therapy in a quickest possible time depending on the degree of the skin burn. Through the utilization Artificial intelligence and advanced technologies, this approach can also provide telemedicine support to diagnose and treat patients remotely, especially in rural areas of least or developing countries where professionals may be scarce. This method can also be utilized in the healthcare facilities where there is limited resources to conduct appropriate clinical diagnosis for detecting skin burn severity. However, this study may be expanded in the future by integrating additional samples, which will allow for the distinction of superficial-partial and deep-partial burns, as well as the estimation of the Total Body Surface Area (TBSA%) of burn for improved clinical assessment for burn patients.

CRediT authorship contribution statement

Sayma Alam Suha: The idea of this article was developed, Background study was conducted, The methodological framework was proposed, The simulation and result analysis was carried out, Part of M.Sc. thesis work, Interpreting the data. **Tahsina Farah Sanam:** The idea of this article was developed, The methodological framework was proposed, Toward rewriting the entire draft article to prepare it for publication, Supervision, Interpreting the data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset generated and analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

The authors in this research are grateful to Dr.Nabil Sharik (MBBS, FCPS, Assistant Surgeon, Sadar, Gopalganj) Dr. Atikur Rahman (MBBS, MO in NiCU & PiCU, M R Khan Shishu Hospital, Bangladesh) and Dr. Fariba Fareen (MBBS, MO, Mymensingh Medical College, Bangladesh) for enriching the study with clinical guidelines.

References

Abubakar, A., Ugail, H., Smith, K. M., Bukar, A. M., & Elmahmudi, A. (2020). Burns depth assessment using deep learning features. *Journal of Medical and Biological Engineering*, *40*(6), 923–933.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., et al. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 1–74.

Arias-Garzón, D., Alzate-Grisales, J. A., Orozco-Arias, S., Arteaga-Arteaga, H. B., Bravo-Ortiz, M. A., Mora-Rubio, A., et al. (2021). COVID-19 detection in X-ray images using convolutional neural networks. *Machine Learning with Applications*, *6*, Article 100138.

Chauhan, J., & Goyal, P. (2020). BPBSAM: Body part-specific burn severity assessment model. *Burns*, *46*(6), 1407–1423.

Crouzet, C., Nguyen, J. Q., Ponticorvo, A., Bernal, N. P., Durkin, A. J., & Choi, B. (2015). Acute discrimination between superficial-partial and deep-partial thickness burns in a preclinical model with laser speckle imaging. *Burns*, *41*(5), 1058–1063.

Dabeer, S., Khan, M. M., & Islam, S. (2019). Cancer diagnosis in histopathological image: CNN based approach. *Informatics in Medicine Unlocked*, *16*, Article 100231.

Hai, T. S., Triet, L., Thai, L., & Thuy, N. (2017). Real time burning image classification using support vector machine. *EAI Endorsed Transactions on Context-Aware Systems and Applications*, *4*(12).

Harjoseputro, Y., Yuda, I., Danukusumo, K. P., et al. (2020). MobileNets: Efficient convolutional neural network for identification of protected birds. *IJASEIT (International Journal on Advanced Science, Engineering and Information Technology)*, *10*(6), 2290–2296.

Herndon, D., Zhang, F., & Lineaweaver, W. (2022). Metabolic responses to severe burn injury. *Annals of Plastic Surgery*, *88*(2), S128–S131.

Jeschke, M. G., van Baar, M. E., Choudhry, M. A., Chung, K. K., Gibran, N. S., & Logsetty, S. (2020). Burn injury. *Nature Reviews Disease Primers*, *6*(1), 1–25.

Jeschke, M. G., Gauglitz, G. G., Kulp, G. A., Finnerty, C. C., Williams, F. N., Kraft, R., et al. (2011). Long-term persistance of the pathophysiologic response to severe burn injury. *PLoS One*, *6*(7), Article e21245.

Kaggle (2022). Skin burn dataset. URL https://www.kaggle.com/shubhambaid/skin-burn-dataset.

Karthik, J., Nath, G. S., & Veena, A. (2021). Deep learning-based approach for skin burn detection with multi-level classification. In *Advances in computing and network communications* (pp. 31–40). Springer.

Keras, T. (2022). Keras documentation: Keras applications-pretrained models. URL https://keras.io/api/applications/.

Kim, H., Jung, W.-K., Park, Y.-C., Lee, J.-W., & Ahn, S.-H. (2022). Broken stitch detection method for sewing operation using CNN feature map and image-processing techniques. *Expert Systems with Applications*, *188*, Article 116014.

Kuan, P., Chua, S., Safawi, E., Wang, H., & Tiong, W. (2017). A comparative study of the classification of skin burn depth in human. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *9*(2–10), 15–23.

Kumari, C. U., Prasad, S. J., & Mounika, G. (2019). Leaf disease detection: feature extraction with K-means clustering and classification with ANN. In *2019 3rd international conference on computing methodologies and communication* (pp. 1095–1098). IEEE.

Laggner, M., Lingitz, M.-T., Copic, D., Direder, M., Klas, K., Bormann, D., et al. (2022). Severity of thermal burn injury is associated with systemic neutrophil activation. *Scientific Reports*, *12*(1), 1–10.

Lee, S., Ye, H., Chittajallu, D., Kruger, U., Boyko, T., Lukan, J. K., et al. (2020). Real-time burn classification using ultrasound imaging. *Scientific Reports*, *10*(1), 1–13.

Lézoray, O., Charrier, C., Cardot, H., & Lefèvre, S. (2008). Machine learning in image processing. *EURASIP Journal on Advances in Signal Processing*, *2008*, 1–2.

Liu, H., Yue, K., Cheng, S., Li, W., & Fu, Z. (2021). A framework for automatic burn image segmentation and burn depth diagnosis using deep learning. *Computational and Mathematical Methods in Medicine*, *2021*.

Mostafa, S., & Wu, F.-X. (2021). Diagnosis of autism spectrum disorder with convolutional autoencoder and structural MRI images. In *Neural engineering techniques for autism spectrum disorder* (pp. 23–38). Elsevier.

Mukti, I. Z., & Biswas, D. (2019). Transfer learning based plant diseases detection using ResNet50. In *2019 4th international conference on electrical information and communication technology* (pp. 1–6). IEEE.

Nandini, G. S., Kumar, A. S., & Chidananda, K. (2021). Dropout technique for image classification based on extreme learning machine. *Global Transitions Proceedings*, *2*(1), 111–116.

Noorbakhsh, S. I., Bonar, E. M., Polinski, R., & Amin, M. S. (2021). Educational case: Burn injury—Pathophysiology, classification, and treatment. *Academic Pathology*, *8*, Article 23742895211057239.

Pabitha, C., & Vanathi, B. (2021). Densemask RCNN: A hybrid model for skin burn image classification and severity grading. *Neural Processing Letters*, *53*(1), 319–337.

Ponticorvo, A., Rowland, R., Baldado, M., Burmeister, D. M., Christy, R. J., Bernal, N. P., et al. (2019). Evaluating clinical observation versus spatial frequency domain imaging (SFDI), laser speckle imaging (LSI) and thermal imaging for the assessment of burn depth. *Burns*, *45*(2), 450–460.

Rangaraju, L. P., Kunapuli, G., Every, D., Ayala, O. D., Ganapathy, P., & Mahadevan-Jansen, A. (2019). Classification of burn injury using Raman spectroscopy and optical coherence tomography: An ex-vivo study on porcine skin. *Burns*, *45*(3), 659–670.

Ren, Z., Qian, K., Dong, F., Dai, Z., Nejdl, W., Yamamoto, Y., et al. (2022). Deep attention-based neural networks for explainable heart sound classification. *Machine Learning with Applications*, Article 100322.

Rice, P. L., & Orgill, D. (2021). Assessment and classification of burn injury. UpToDate, [Internet].

Rowland, R. A., Ponticorvo, A., Baldado, M. L., Kennedy, G. T., Burmeister, D. M., Christy, R. J., et al. (2019). Burn wound classification model using spatial frequency-domain imaging and machine learning. *Journal of Biomedical Optics*, *24*(5), Article 056007.

Şevik, U., Karakullukçu, E., Berber, T., Akbaş, Y., & Türkyılmaz, S. (2019). Automatic classification of skin burn colour images using texture-based feature extraction. *IET Image Processing*, *13*(11), 2018–2028.

Sharma, O. (2019). A new activation function for deep neural network. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing* (pp. 84–86). IEEE.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1–48.

Shpichka, A., Butnaru, D., Bezrukov, E. A., Sukhanov, R. B., Atala, A., Burdukovskii, V., et al. (2019). Skin tissue regeneration for burn injury. *Stem Cell Research & Therapy*, *10*(1), 1–16.

Stoica, A. E., Chircov, C., & Grumezescu, A. M. (2020). Hydrogel dressings for the treatment of burn wounds: an up-to-date overview. *Materials*, *13*(12), 2853.

Tammina, S. (2019). Transfer learning using VGG-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(10), 143–150.

Tchito Tchapga, C., Mih, T. A., Tchagna Kouanou, A., Fozin Fonzin, T., Kuetche Fogang, P., Mezatio, B. A., et al. (2021). Biomedical image classification in a big data architecture using machine learning algorithms. *Journal of Healthcare Engineering*, *2021*.

Tran, H. S., Le, T. H., & Nguyen, T. T. (2016). The degree of skin burns images recognition using convolutional neural network. *Indian Journal of Science*, *9*(45), 1–6.

Wang, Y., Beekman, J., Hew, J., Jackson, S., Issler-Fisher, A. C., Parungao, R., et al. (2018). Burn injury: Challenges and advances in burn wound healing, infection, pain and scarring. *Advanced Drug Delivery Reviews*, *123*, 3–17.

Yadav, D., Sharma, A., Singh, M., & Goyal, A. (2019). Feature extraction based machine learning for human burn diagnosis from burn images. *IEEE Journal of Translational Engineering in Health and Medicine*, *7*, 1–7.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights Into Imaging*, *9*(4), 611–629.

Zhu, Y., & Huang, C. (2012). An improved median filtering algorithm for image noise reduction. *Physics Procedia*, *25*, 609–616.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, *109*(1), 43–76.