

Article

Comparison of Deep Transfer Learning Techniques in Human Skin Burns Discrimination

Aliyu Abubakar ^{1,2,*} , Mohammed Ajuji ² and Ibrahim Usman Yahya ²¹ Centre for Visual Computing, Faculty of Engineering and Informatics, University of Bradford, Bradford BD7 1DP, UK² Department of Computer Science, Faculty of Science, Gombe State University, Tudun Wada 760231, Nigeria; majuji@gsu.edu.ng (M.A.); usmanitgombe@gsu.edu.ng (I.U.Y.)

* Correspondence: a.abubakar6@bradford.ac.uk; Tel.: +44-7-574-758-316

Received: 11 March 2020; Accepted: 7 April 2020; Published: 11 April 2020



Abstract: While visual assessment is the standard technique for burn evaluation, computer-aided diagnosis is increasingly sought due to high number of incidences globally. Patients are increasingly facing challenges which are not limited to shortage of experienced clinicians, lack of accessibility to healthcare facilities and high diagnostic cost. Certain number of studies were proposed in discriminating burn and healthy skin using machine learning leaving a huge and important gap unaddressed; whether burns and related skin injuries can be effectively discriminated using machine learning techniques. Therefore, we specifically use transfer learning by leveraging pre-trained deep learning models due to deficient dataset in this paper, to discriminate two classes of skin injuries—burnt skin and injured skin. Experiments were extensively conducted using three state-of-the-art pre-trained deep learning models that includes ResNet50, ResNet101 and ResNet152 for image patterns extraction via two transfer learning strategies—fine-tuning approach where dense and classification layers were modified and trained with features extracted by base layers and in the second approach support vector machine (SVM) was used to replace top-layers of the pre-trained models, trained using off-the-shelf features from the base layers. Our proposed approach records near perfect classification accuracy in categorizing burnt skin and injured skin of approximately 99.9%.

Keywords: burns; pressure ulcer; bruises; deep learning; classification

1. Introduction

Burns are injuries affecting both elderly, young and children, causing loss of lives and subjecting both families and nations to additional economic burden. An estimated sum of \$7.5 billion has been reported as the annual expenditure for the management of burn injuries in America [1], third most devastating cause of unintentional injuries and second most serious injury affecting elderly people. Another report by Reference [2] shows that burn is the most common injury affecting children below the age of five in England, causing about 7000 admissions in hospitals each year. Many research reports [3–5] have lamented a worrisome situation in poor countries where high cases (up to 90%) of burns injuries are recorded with high percentage rate of mortality and was attributed to lack of social awareness programs related to prevention and safety measures, while most of the cost-effective and safety measures are practically implemented in high income countries [1]. In general, studies [6,7] have shown that about 265,000 deaths are reported every year, as such burns have become eminent health concern globally.

Burns heal within a three-week window period though superficial burn heals in 5–7 days post-burn. Effective and early assessment is important in order to determine whether a conservative treatment or surgical excision is required to avoid further complications. Assessment of burns is effective clinically

when the burn injuries involves superficial or full-thickness and requires experienced dermatologist or surgeons and time consuming [1]. Overall, the assessment of burns by experienced clinicians is accurate in only between 65–75% [8,9]. For this reason, certain number of modalities were introduced to aid clinicians in making accurate evaluation of the burn injuries. Laser Doppler Imaging (LDI) is one of the examples of the introduced modalities that received unanimous acceptance and approved by burn community. LDI operates based on Doppler effect principle and assesses burn depth by providing healing potentials of the wounds. The advantages of using LDI includes its non-invasive nature and effectiveness than human experts. However, LDI has some shortfall such as high affordability cost, assessment ineffectiveness in first 24 h after burn [10] and not suitable for pediatric patients due to movement artifacts.

Interestingly, the use of machine learning techniques has been penetrating different sectors including healthcare to boost operational efficiency and help minimize inconsistencies due to human errors. Therefore, in this paper we proposed the use of machine learning to discriminate burns and injured skin. The motivation behind this proposal are—(1) human expertise is not readily available in most remote locations; (2) assessment by human experts is time consuming and subjective [11]; (3) up-to-date approved device (i.e., LDI) for burns assessment by burn community is highly expensive [12] and (4) to the best of our knowledge, no technique was proposed to address discriminating burns and injured skin, as such we proposed the use of machine learning to discriminate whether a given image is burn or injured. The use of machine learning can serve as alternative to human expertise thereby aiding in timely decision making. Transfer learning technique was used due to deficient datasets to train a model from scratch. The rest of the paper is organized as follows—Section 2 provides a review related literature; Section 3 presents materials used and the methodology; Section 4 presents results and discussion; while in Section 5 conclusion was presented.

2. Literature Review

Deep learning evolved from a tiny architecture called Artificial Neural Network (ANN), popularly known for classification task which was inspired by human brain. Perceptron is a basic unit of ANN which has an associated weight, bias and then a manipulative function applied to supplied input to obtain a desired outcome. During the process, both weights and biases are adjusted regularly until a convincing or satisfactory output is generated by the algorithm. This process of adjusting weights and biases is called Back-propagation. The introduction of non-linearity function made neural network more successful and deeper enough to extract complex patterns from the data. The deeper version of ANN is called deep neural network with many variations such as Convolutional Neural Network (CNN). In CNN, convolution layers are equipped with filters or kernels that slide over a given image to extract useful patterns and the output of the sliding operation is known as feature map which is then made available as input to next layer. However, the success of deep learning algorithms lies with the availability of enough training data and powerful computational machines. CNN have achieved state-of-the-art recognition accuracies in many classification problems including plant disease detection [13], cancer [14,15] and skin burns assessment [9].

A study in Reference [9] to discriminate burns and healthy skin using off-the-shelf features was proposed. The study utilized a deep pre-trained neural network model to extract patterns from the given images. This was done by passing the images through the base convolution layers of the adopted model without the top classification layer, following pre-processing steps such normalization and resizing processes. Support Vector Machine (SVM) was used as replacement of the classification layer and was trained using feature scores obtained from the convolution operations. 10-folds cross-validation strategy was applied during SVM training and interestingly, 99.5% accuracy was achieved in assessing whether a given image is burnt or healthy.

Skin burns are categorized into classes depending on whether the injury affects the top-most layer, intermediate or deep layer. In this regard, the authors of Reference [16] proposed a study to classify healthy skin, superficial and deep dermal (full-thickness) burns using pre-trained deep learning

model and SVM. Specifically, they used ResNet101 for the extraction of image feature extraction and SVM for the classification task, in which identification accuracy of 99.9% was achieved using 10-folds cross-validation strategy.

In similar development, research by the authors of Reference [10] use of deep learning models to assess whether a given image is superficial burn, superficial to intermediate partial thickness burns, intermediate to deep partial thickness burns, deep partial and full thickness burns, healthy skin or background image. Four pre-trained deep learning models were used to classify the burn images transfer learning approach, in which earlier convolution layers were frozen to extract features and the top-most layers were modified and trained using features obtained from the frozen layers. This process was applied to several deep neural network models including VGG-16 which recorded 86.7% accuracy, 79.4% for GoogleNet, 82.4% for ResNet-50 and 88.1% for ResNet-101.

Studies in Reference [17] proposed the use of off-the-shelf features with SVM classifier to classify burns in both Caucasian and African patients. Off-the-shelf features were extracted using three pre-trained deep learning models which include VGG-16, VGG-19 and VGG-Face while in each case SVM was used as the classifier. Classification accuracy on Caucasian images, VGG-16, VGG-19 and VGG-Face achieved 99.3% 98.3% and 96.3% respectively. Using African images, VGG-16, VGG-19 and VGG-Face achieved 98.9%, 97.5% and 97.2% respectively. Classification accuracy of 98.6%, 97.6% and 95.2% were achieved by VGG-15, VGG-19 and VGG-Face respectively when both Caucasian and African datasets were combined.

Pressure Ulcer and Bruises

Pressure ulcer injuries are mostly sustained on body locations at bony prominence coupled with contraction of the skin/tissue with external hard surface. According to study in UK over 700,000 people get affected by pressure ulcer every year, costing health service more than £3.8 million per day and estimated sum of £1.4–£2.4 billion per year [18] These injuries are developed by elderly individuals, people with physical disabilities and patients with severe illnesses whose mobility is very limited in most cases. Pressure ulcer can prolong hospital-stay, especially people with severe burn wounds are at risk of developing this type of injury due to limited mobility and in extreme cases may lead to increase in mortality rate and quality of life reduction. Bruises (contusions) are common in maltreated children and become noticeable via careful examinations by clinicians. Contusions are also sustained accidentally in toddlers due to tumbling plays during their early exploration pursuit. Similar reports related burn injuries have indicated its prominence in children [19,20]. Those contusions occurring at bony prominences such as knees, elbows, anterior tibiae and forehead are mostly due to accidental causes [21]. Deviation from those areas of bony prominence raises suspicion of possible child maltreatment. Likewise, accidental patterns of contusions in disabled people are found on feet, hands, arms and abdomen mostly attributed to the use of such body parts as mobility aid but not often well known on the neck, genitals, chest and ears.

In this paper, we address the discrimination of burns and injured skin that share similar characteristics. This is aim in ensuring that the decision-making is quick and accurate treatment that relates to a specific injury is provided without subjecting patients to further undesirable complications

3. Materials and Methods

The main concept here is to classify skin burn images and other skin injuries that have close or similar physiological appearances using transfer learning concept. Transfer learning is the use of pre-trained convolutional neural network model as a feature extractor in which the features extracted are then used to train a new layer(s) or classification algorithm such as support vector machine. Therefore, we used two different methods of transfer learning to discriminate between burns and other tissue injuries. Firstly, we used fine-tuning strategy as described in Section 3.3.1 and secondly used off-the-shelf features as presented in Section 3.3.2

3.1. Dataset

Our database contains two sets of images—skin burn images and injured skin (bruised and pressure ulcer) images. This means what we have is a binary classification problem consisting mainly Caucasian participants as depicted in Figure 1. The burn images were ethically acquired from Bradford hospital in United Kingdom whereas both skin bruises and pressure ulcer wounds were acquired via the use of internet search [22]. The database contains 300 burn images and 250 bruised and pressure ulcer wound images and for ease of references, we denoted the combination of both pressure ulcer and bruise images as PUB.



Figure 1. Sample of dataset images.

3.2. Data Pre-Processing

Pre-processing is a crucial step to ensure unwanted noise is minimized in order to enhance computational efficiency. Unnecessary background in the images were cropped out manually prior to feeding the images into the deep learning model while normalization was conducted automatically during the run-time as well as artificially enlarging the size of the database (known as augmentation) via the use of different transformation processes such as rotation and flipping, resulting to a larger database of 1160 images in each category.

3.3. Feature Extraction

Deep learning models are greedy algorithms that require huge data to learn effectively. Having such data was a big problem to the research community in the field, in addition to lack of powerful computational hardware. With the emergence of large databases such as ImageNet [23] and hardware-accelerated devices such as graphical processing unit (GPU) respectively, deep learning research was revolutionized and the world started witnessing huge deep learning algorithms such as AlexNet,

VGGNet, GoogleNet and Residual Networks (ResNet) to mention but few. AlexNet the pioneer deepest learning model to be trained on ImageNet database to categorize 1000 different objects in 2012 [24] made an impressive breakthrough. Thereafter, different deeper neural models than AlexNet were proposed by different research communities including Visual Geometry Group (VGG) from Oxford University and GoogleNet from Google both in 2014 [25]. The deeper the network, the more patterns it can extract from the given data. However, building such deep neural network models have faced a vanishing gradient problem and difficulty during backpropagation and optimization processes. Microsoft Research Asia (MRA) in 2015 [26] proposed a very deep neural network model called Residual Networks (ResNet) to ensure that the performance of top layers is as good as the lower layers without vanishing gradient and optimization problem. This was achieved by using identity connection in parallel to the main convolution layers as illustrated in Figure 2, making back-propagation easier, quick learning and most importantly allows building network deeper enough to extract complex patterns from the data while maintaining an optimum accuracy improvement. MRA succeeded in developing very deep neural networks of at least fifty convolution layers such as ResNet50, ResNet101 and ResNet152.

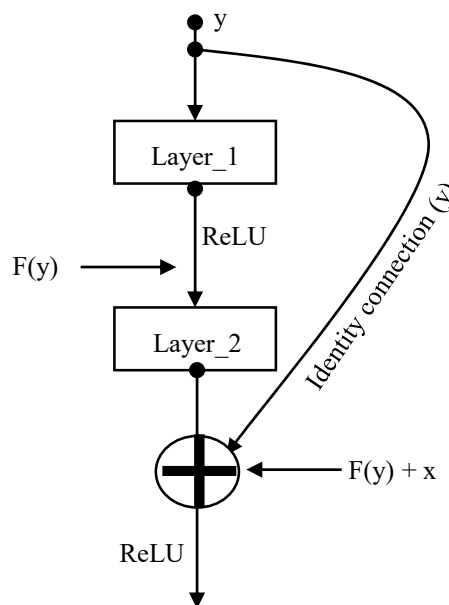


Figure 2. Illustration of ResNet identity connection.

Interestingly, those pre-trained models can be used to achieve desired outcome in areas with deficient data via a process called transfer learning [27]. Transfer learning process is used in two approaches—fine-tuning where some modifications are made and as an off-the-shelf feature extractor where features are extracted in order to train a machine learning classifier. Both transfer learning approaches were utilized in this paper using the three ImageNet pre-trained models (that is, ResNet50, ResNet101 and ResNet152) to automatically discriminate whether a given abnormal skin image is burn or PUB.

3.3.1. Fine-Tuning

In this approach, top-most layers (that is, fully connected and the classification layers) were removed, two new dense layers and a classification layer with a sigmoid function were used to replace the removed layers. These new added layers were trained using features extracted from the transferred learned layers of the pre-trained deep learning models. Note that the transferred layers were frozen to only extract image patterns without adjusting their weights. This scenario is applied to all the three pre-trained ResNet models as illustrated in Figure 3 where 80% of the datasets were used for training

and the 20% of the datasets were reserved for validation. None of the instances in the validation dataset featured in the training dataset.

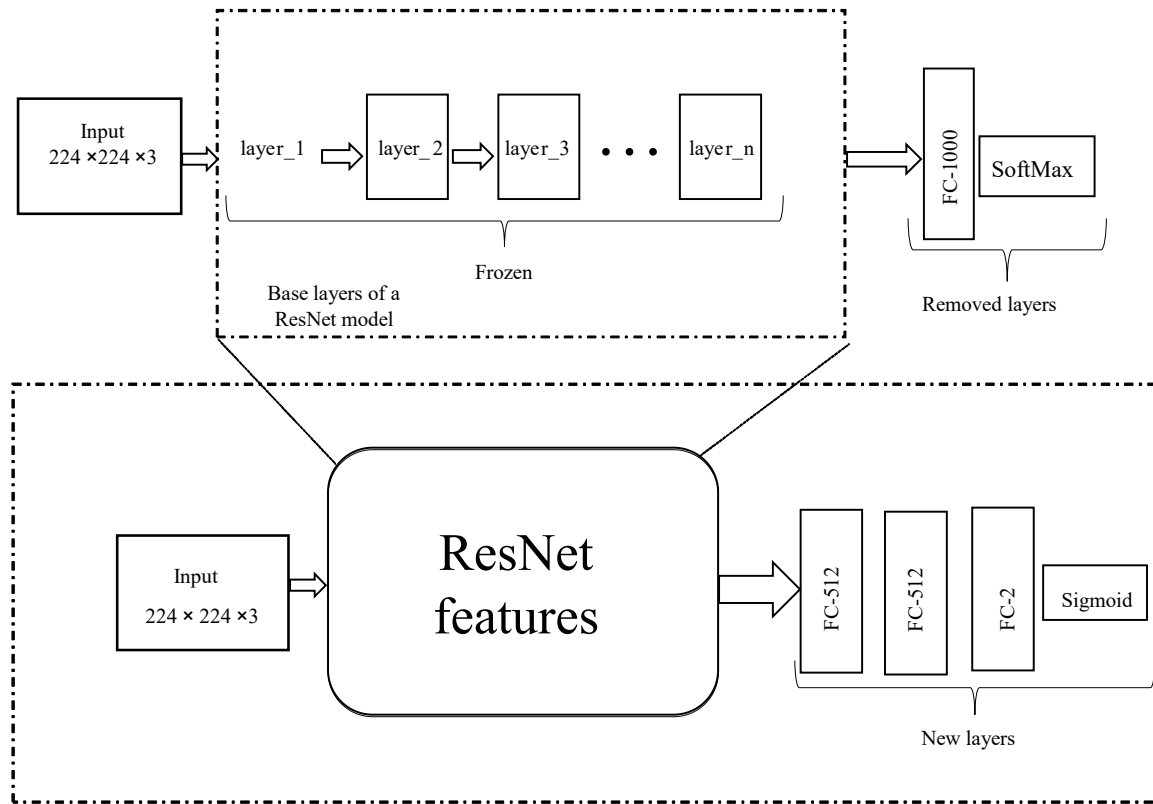


Figure 3. Illustration of fine-tuning.

3.3.2. Off-the-Shelf Feature Extraction

Here, base layers of a pre-trained deep learning model are used to extract useful patterns from the data. These patterns or features are called off-the-shelf features which are then transferred to train machine learning algorithm such as SVM and decision tree. Similarly, base layers of the pre-trained model were frozen to extract useful patterns from the data as illustrated in Figure 4. We used linear SVM to classify extracted features into whether the features belong to Burns or PUB. SVM is a supervised machine learning algorithm used for both classification and regression purposes. The SVM is trained to find the most optimum separating hyperplane that discriminates the two classes of images by solving optimization problem.

Given a training set,

$$\{x_i, y_i\}_{1 \leq i \leq n}, x_i \in R^d, y_i \in \{+1, -1\}. \quad (1)$$

SVM finds an optimum separating hyperplane that discriminate the two classes by solving

$$\begin{cases} \min & \frac{1}{2} \|w\|^2 \\ w, b & \\ \text{with} & y_i \cdot (w \cdot x + b) \geq 1 \end{cases} \quad (2)$$

for $i = 1, \dots, n$ are the observations, w and b are weights and biases learned during the training respectively. The SVM learned such that $y_i \in \{+1, -1\}$, where burn images are denoted by $+1$ and PUB images are denoted by -1 . Thus, the optimum separating hyperplane was computed which discriminated Burns and PUB into their respective category. A K -fold cross validation technique was used in training the SVM classifier, where a value of 10 was chosen K .

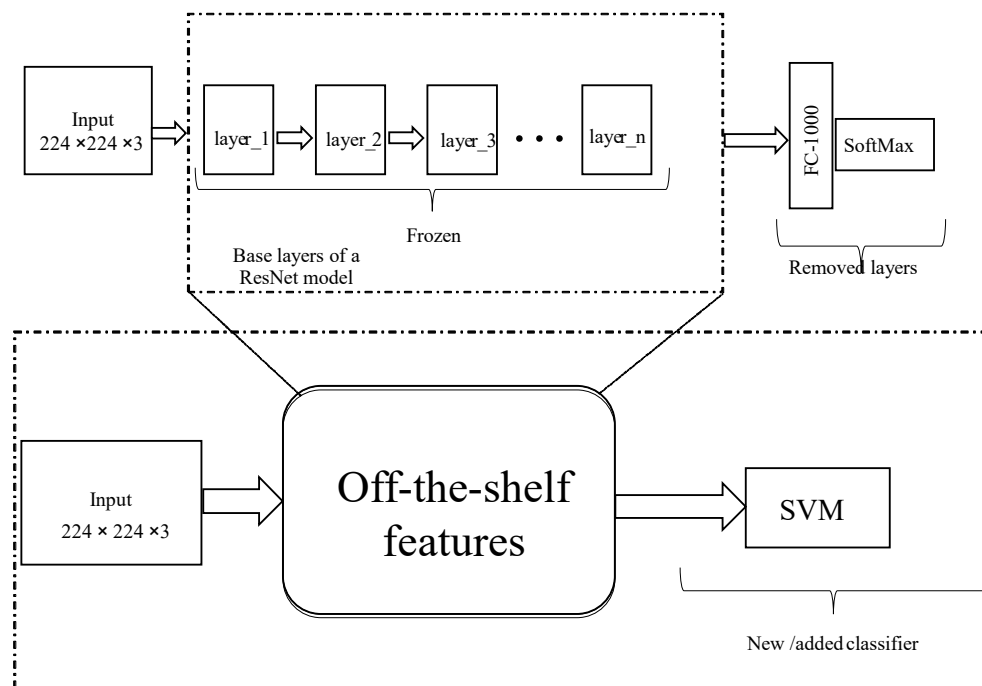


Figure 4. Illustration of off-the-shelf feature extraction and classification.

4. Results

During the experiment process using fine-tuning approach, 78% of the data in each class was used for training the algorithm while the reserved 22% was used for validation/testing. High proportion of the data used in the training process was to ensure the algorithm learns more representation patterns from the data, we also ensured that no instances used during training process were also present in the validation set.

Another important decision we made is the use of computational resource from google, called Google Collab also known Collaboratory [28]. This resource provides free access to on-demand GPU hardware-accelerator.

Figures 5–7 depicted the training processes of fine-tuned ResNet50, ResNet101 and ResNet152 respectively, using 200 epochs in each. Figure 8 shows training and validation accuracies using ResNet50, in which the result shows a near perfect classification output, while the result depicted in Figure 9 provides training and validation losses of the fine-tuned ResNet50.

```
Epoch 192/200
56/56 [=====] - 8s 145ms/step - loss: 0.0051 - acc: 0.9976 - val_loss: 0.1595 - val_acc: 0.9810
Epoch 193/200
56/56 [=====] - 8s 146ms/step - loss: 0.0141 - acc: 0.9988 - val_loss: 0.1230 - val_acc: 0.9786
Epoch 194/200
56/56 [=====] - 8s 148ms/step - loss: 0.0257 - acc: 0.9952 - val_loss: 0.1329 - val_acc: 0.9786
Epoch 195/200
56/56 [=====] - 8s 145ms/step - loss: 0.0066 - acc: 0.9976 - val_loss: 0.2271 - val_acc: 0.9786
Epoch 196/200
56/56 [=====] - 8s 145ms/step - loss: 0.0178 - acc: 0.9958 - val_loss: 0.1591 - val_acc: 0.9833
Epoch 197/200
56/56 [=====] - 8s 150ms/step - loss: 0.0245 - acc: 0.9952 - val_loss: 0.1538 - val_acc: 0.9810
Epoch 198/200
56/56 [=====] - 8s 146ms/step - loss: 0.0020 - acc: 0.9988 - val_loss: 0.2219 - val_acc: 0.9810
Epoch 199/200
56/56 [=====] - 8s 143ms/step - loss: 0.0054 - acc: 0.9988 - val_loss: 0.1447 - val_acc: 0.9833
Epoch 200/200
56/56 [=====] - 8s 146ms/step - loss: 0.0067 - acc: 0.9976 - val_loss: 0.1570 - val_acc: 0.9810
```

Figure 5. Showing ResNet50 training process using fine-tuning.

```

Epoch 171/200
56/56 [=====] - 10s 180ms/step - loss: 0.0062 - acc: 0.9976 - val_loss: 0.0777 - val_acc: 0.9857
Epoch 172/200
56/56 [=====] - 10s 179ms/step - loss: 0.0011 - acc: 0.9994 - val_loss: 0.0693 - val_acc: 0.9857
Epoch 173/200
56/56 [=====] - 10s 177ms/step - loss: 0.0031 - acc: 0.9982 - val_loss: 0.0679 - val_acc: 0.9881
Epoch 174/200
56/56 [=====] - 10s 177ms/step - loss: 3.7694e-04 - acc: 1.0000 - val_loss: 0.0716 - val_acc: 0.9881
Epoch 175/200
56/56 [=====] - 10s 176ms/step - loss: 1.4302e-04 - acc: 1.0000 - val_loss: 0.0722 - val_acc: 0.9881
Epoch 176/200
56/56 [=====] - 10s 177ms/step - loss: 1.7204e-04 - acc: 1.0000 - val_loss: 0.0765 - val_acc: 0.9881
Epoch 177/200
56/56 [=====] - 10s 179ms/step - loss: 1.6121e-04 - acc: 1.0000 - val_loss: 0.0742 - val_acc: 0.9881
Epoch 178/200
56/56 [=====] - 10s 176ms/step - loss: 0.0018 - acc: 0.9988 - val_loss: 0.0851 - val_acc: 0.9881

```

Figure 6. Showing ResNet101 training process using fine-tuning.

```

Epoch 173/200
56/56 [=====] - 13s 234ms/step - loss: 0.0139 - acc: 0.9964 - val_loss: 0.0233 - val_acc: 0.9952
Epoch 174/200
56/56 [=====] - 13s 232ms/step - loss: 0.0089 - acc: 0.9970 - val_loss: 0.0598 - val_acc: 0.9905
Epoch 175/200
56/56 [=====] - 13s 233ms/step - loss: 0.0063 - acc: 0.9976 - val_loss: 0.0195 - val_acc: 0.9929
Epoch 176/200
56/56 [=====] - 13s 230ms/step - loss: 0.0046 - acc: 0.9994 - val_loss: 0.0195 - val_acc: 0.9905
Epoch 177/200
56/56 [=====] - 13s 230ms/step - loss: 0.0053 - acc: 0.9988 - val_loss: 0.0202 - val_acc: 0.9905
Epoch 178/200
56/56 [=====] - 13s 232ms/step - loss: 0.0141 - acc: 0.9976 - val_loss: 0.0376 - val_acc: 0.9929
Epoch 179/200
56/56 [=====] - 13s 229ms/step - loss: 0.0232 - acc: 0.9952 - val_loss: 0.0818 - val_acc: 0.9905
Epoch 180/200
56/56 [=====] - 13s 234ms/step - loss: 0.0145 - acc: 0.9982 - val_loss: 0.0106 - val_acc: 0.9952
Epoch 181/200
56/56 [=====] - 13s 233ms/step - loss: 0.0057 - acc: 0.9976 - val_loss: 0.0254 - val_acc: 0.9929
Epoch 182/200
56/56 [=====] - 13s 234ms/step - loss: 0.0015 - acc: 0.9994 - val_loss: 0.0160 - val_acc: 0.9929
Epoch 183/200
56/56 [=====] - 13s 232ms/step - loss: 0.0032 - acc: 0.9994 - val_loss: 0.0158 - val_acc: 0.9929
Epoch 184/200
56/56 [=====] - 13s 233ms/step - loss: 0.0137 - acc: 0.9970 - val_loss: 0.0133 - val_acc: 0.9929
Epoch 185/200
56/56 [=====] - 13s 233ms/step - loss: 0.0166 - acc: 0.9958 - val_loss: 0.0618 - val_acc: 0.9929

```

Figure 7. Showing ResNet152 training process using fine-tuning.

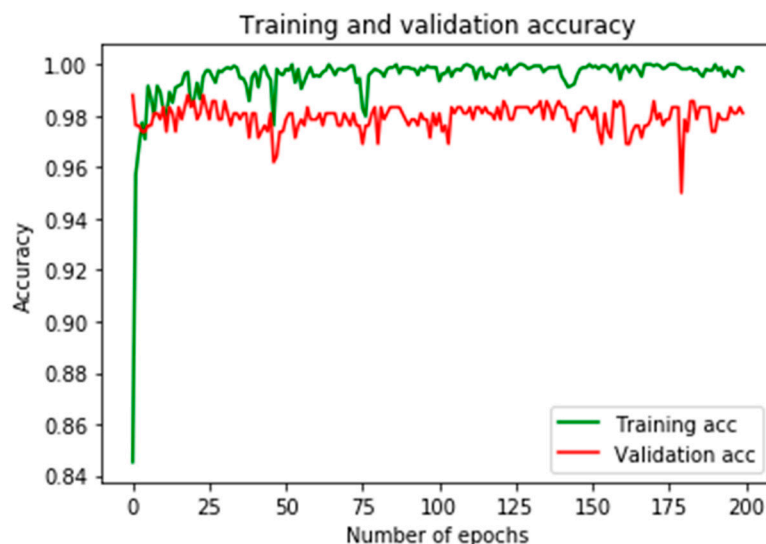


Figure 8. Showing training and validation accuracies using ResNet50.

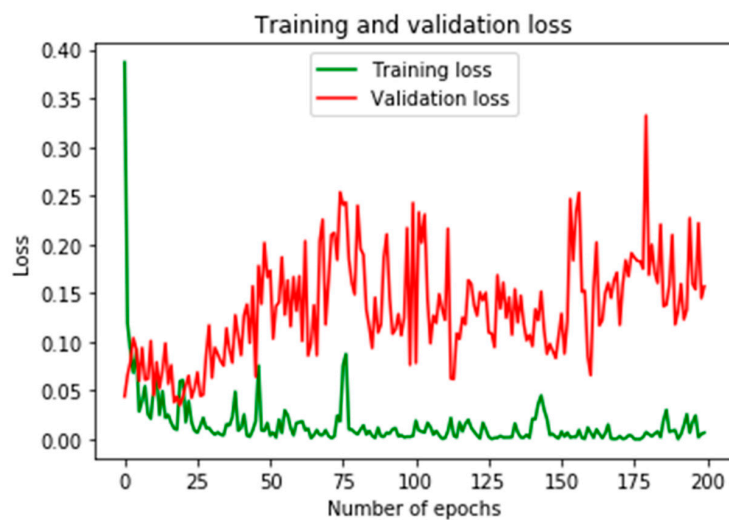


Figure 9. Showing training and validation losses using ResNet50.

4.1. Results from Fine-Tuning Strategy

Here, we present the results obtained using fine-tuning approach where 80% of the datasets was used for trained and the remaining 20% was used for validation. High proportion of the dataset was allocated for training because the model requires to learn from more data, this enable learning more pattern representation from the data, while the validation data is used to assess the trained model's efficacy.

For ResNet101, training and validation accuracies were shown in Figure 10. Figure 11 provides the results of training and validation losses. Here, we observed more stability in terms of accuracies and losses than the previous results from ResNet50.

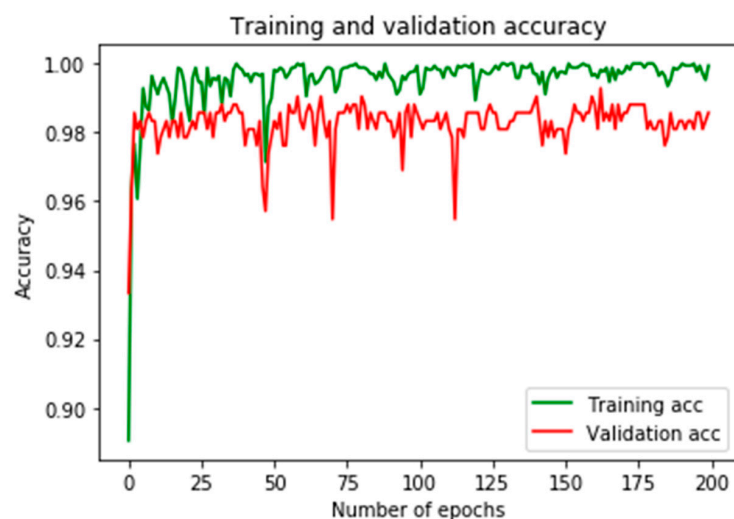


Figure 10. Showing training and validation accuracies using ResNet101.

Lastly, both training and validation accuracies using ResNet152 as depicted in Figure 12 has surpassed what was obtained in both ResNet50 and ResNet101. Similarly, training and validation losses as depicted in Figure 13 show an impressive classification performance. This has proven a point that the deeper the network is; the more patterns the network is capable of extracting from the data.

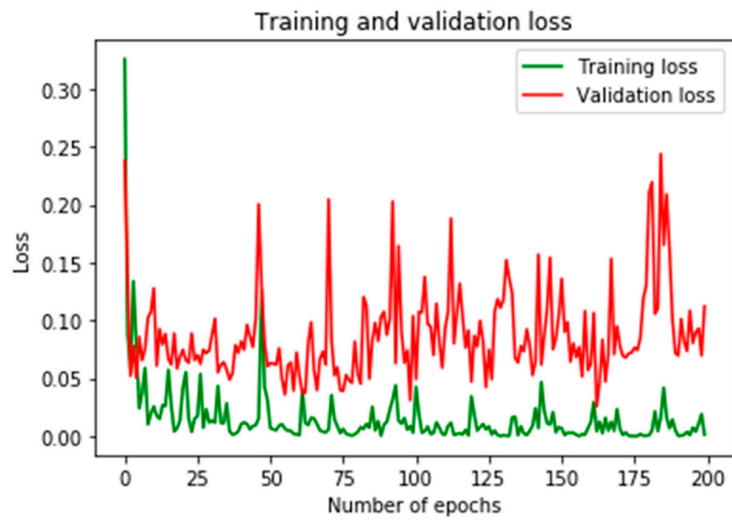


Figure 11. Showing training and validation losses using ResNet101.

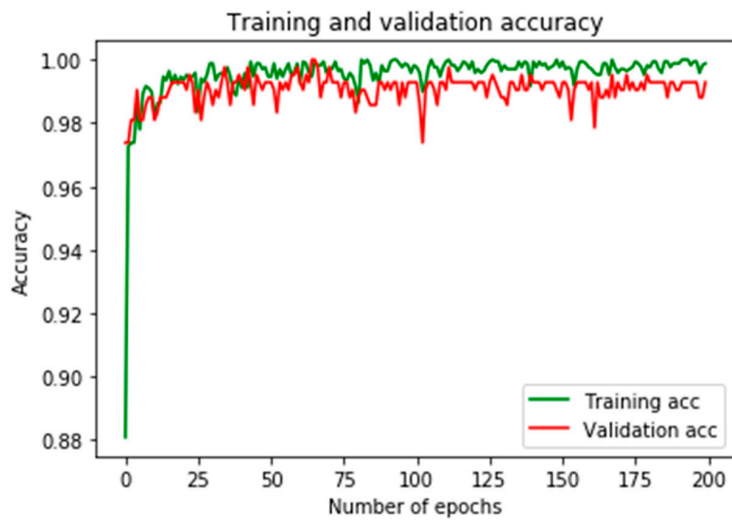


Figure 12. Showing training and validation accuracies using ResNet152.

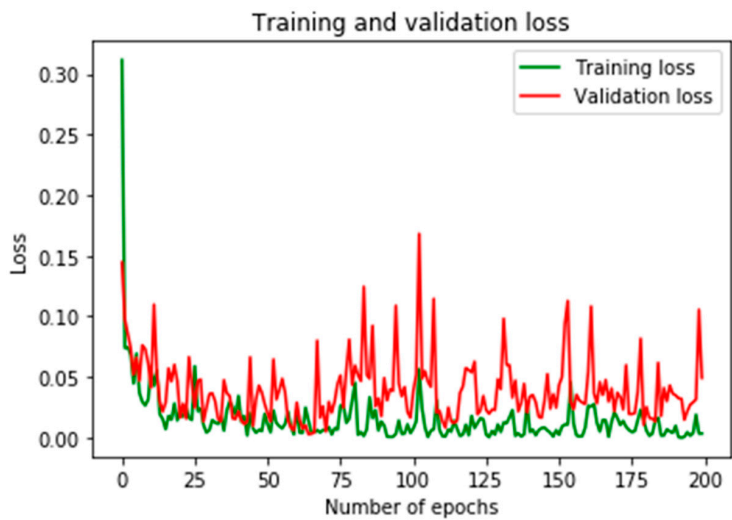


Figure 13. Showing training and validation losses using ResNet152.

4.2. Results from Off-the-Shelf features and SVM

Using off-the-shelf features and SVM classifier, a near perfect results were achieved using features in each of the three pre-trained models. The classification results were presented in Table 1 for ResNet50 features, Table 2 for ResNet101 features and Table 3 for ResNet152 features.

Table 1. Support Vector Machine (SVM) classification using ResNet50 features.

Output Classes	Target Classes	
	PUB	Burns
PUB	1160	3
Burns	0	1157

Table 2. SVM classification using ResNet101 features.

Output Classes	Target Classes	
	PUB	Burns
PUB	1160	3
Burns	0	1157

Table 3. SVM classification using ResNet152 features.

Output Classes	Target Classes	
	PUB	Burns
PUB	1160	1
Burns	0	1159

The sensitivity and specificity of the classification output were calculated. Sensitivity (true positive rate) is the proportion of actual burn images that are correctly classified while specificity is the proportion of actual non burn images that are correctly classified [9]. Sensitivity is determined by the following mathematical equation— $\text{sensitivity} = \frac{TP}{TP+FN} \times 100$; specificity is determined by the following mathematical equation; $\text{specificity} = \frac{TN}{TN+FP} \times 100$. In each of the Tables 1–3 above, target (actual) classes are presented as columns while rows of the tables represent output (predicted) classes. The intersection where Burns in the second column and Burns in the second-row meet represent the true positives (TP); the intersection where PUB in the first column and PUB in the first-row meet represents true negatives (TN). FP stands for false positive which means PUB that were predicted by the algorithm as Burns which can be found at first cell in the second row (second cell in the second column). FN stands for false negative which means Burns that were predicted as by the algorithm as PUB which can be found at second cell of the first row (first cell of the second column). Table 4 presents the sensitivity and specificity of the SVM classifier for each of the features in the three models.

Classification accuracy is the ratio of all correct predictions to the total number of all the samples used which can be calculated by the following mathematical equation— $\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100$. Table 5 shows the classification accuracies obtained by SVM classifier on all the feature sets. Near perfect results were obtained in each case, with best recognition accuracy using ResNet152 features in each of the approaches used. However, classification is more efficient using ResNet152 features which took approximately 1660 s to complete compared to 2116 and 2705 s for both ResNet101 and ResNet152 respectively. With off-the-shelf features from all the three feature sets, SVM performance is more impressive with ResNet152 and more efficient than the other two features from ResNet50 and ResNet152 as presented in Table 5.

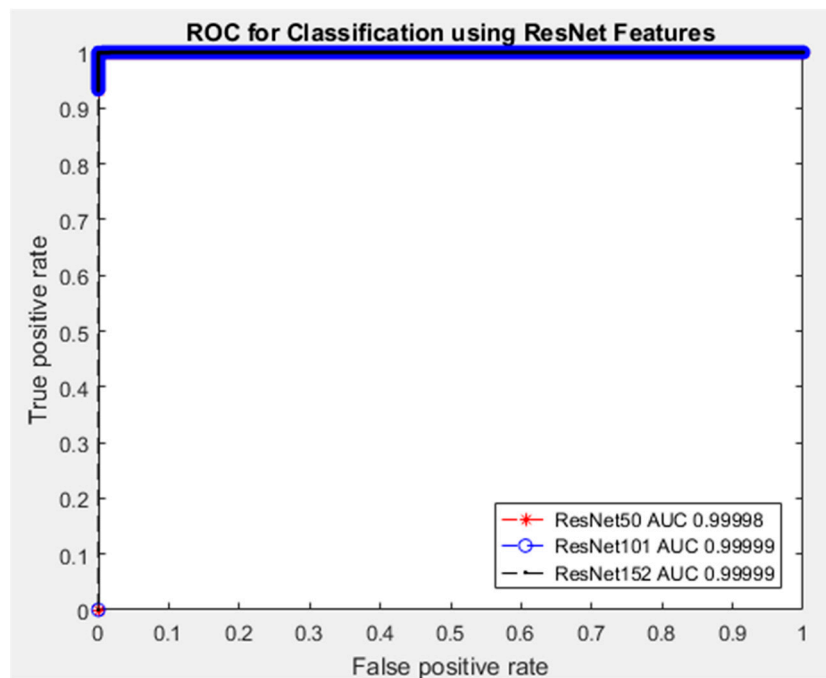
Table 4. Performance Metrics.

Models	Sensitivity	Specificity
ResNet50	0.9974	1.0000
ResNet101	0.9974	1.0000
ResNet152	0.9991	1.0000

Table 5. Classification Accuracies.

Models	Fine-Tuning		SVM	
	Accuracy	Time (s)	Accuracy	Time (s)
ResNet50	0.9833	1659.8	0.9987	15.2
ResNet101	0.9881	2116.1	0.9987	14.6
ResNet152	0.9952	2704.6	0.9996	13.0

Furthermore, the Receiver Operating Characteristics (ROC) curve is another important tool for analyzing the diagnostic performance of a binary classifier. ROC curve gives the trade-off between sensitivity and specificity as the threshold varies [29]. One important feature of ROC curve is the Area Under the Curve (AUC) which provides the overall summary of the classifier's performance. The AUC ranges from 0.5 to 1, where AUC of 0.5 or less indicates poor diagnosis (meaningless assessment) while AUC of 1 indicates perfect assessment. Figure 14 shows the ROC curve and the corresponding AUC values of the SVM classifier for each of the three sets of features. Near perfect results with approximately AUC = 0.99 in all scenarios. What these AUC values indicate is that, a chosen randomly burn images will have a higher test value than a chosen PUB image, which means higher AUC value is indicative of the burn.

**Figure 14.** Receiver Operating Characteristic (ROC) curve for the binary classification of Burns and PUB.

Moreover, though studies in the literature were mainly based on binary discrimination between burn and healthy skin, comparing our finding as presented in Table 6 shows impressive outcome despite addressing most complicated issue (burns and other skin injuries).

Table 6. Comparison with the state-of-the-art findings.

Features and Classifier	Datasets	Classification Accuracy
ResNet101 + SVM [9]	Burns & Healthy Skin (Caucasian)	99.5%
VGG-16 + SVM [17]	Burns & Healthy Skin (Caucasian)	99.3%
VGG-19 + SVM [17]	Burns & Healthy Skin (Caucasian)	98.3%
VGG-Face + SVM [17]	Burns & Healthy Skin (Caucasian)	96.3%
ResNet152 + SVM	Burns & PUB (Caucasian)	99.9%

5. Conclusions

This paper provides the use of deep transfer learning (TL) techniques for burns discrimination using the state-of-the-art deep convolutional neural networks via the use of two TL approaches. In the first TL approach, the top-most layer and the classification were removed, and new added layers were trained using image features from the remaining convolution layers. In the second technique, similarly top-most layer and the classification layer were removed but instead of adding new layers, SVM was used as a substitute for classification purpose. In each of the two TL approaches, three pre-trained deep convolutional neural networks were used for the pattern recognition from the given images (burns and PUB).

The results in each approach and from each model show a near perfect recognition accuracy, where in the first TL approach ResNet50 achieved up to 98.3% accuracy in 1659 s, ResNet101 achieved 98.8% accuracy in 2116 s while resnet152 achieved 99.5% accuracy in 2704 s. Using Resnet152 shows most outstanding performance compared to other two models but ResNet50 appears to be most efficient in terms of timely assessment. This is due to variation of the model's complexity such as network sizes and the parameters that need to be optimized. In the second TL approach, near perfect recognition accuracy were obtained with outstanding performance from ResNet152 features. SVM recorded 99.9% accuracy in 13 s using ResNet152 features and 99.8% accuracy in 14 and 15 s for both ResNet101 and ResNet50 features respectively.

From the performances recorded, conclusion is made that using pre-trained models as feature extractors while SVM as the classifier shows to be the most appropriate TL technique in a situation with deficient datasets. However, the limitation of our study includes discrimination of burns as it relates to severity known as depth is not addressed here, this is intended to be addressed in future work.

Author Contributions: Conceptualization, A.A.; methodology, A.A.; software, A.A.; formal analysis, A.A.; investigation, A.A.; resources, A.A. and M.A.; data curation, A.A., M.A. and I.U.Y.; writing—original draft preparation, A.A.; writing—review and editing, I.U.Y.; supervision, A.A.; funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Petroleum Technology Development Fund (PTDF) Nigeria, grant number PTDF/ED/OSS/PHD/AA/1104/17.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dexter, G.; Patil, S.; Singh, K.; Marano, M.A.; Lee, R.; Petrone, S.J.; Chamberlain, R.S. Clinical outcomes after burns in elderly patients over 70 years: A 17-year retrospective analysis. *Burns* **2018**, *44*, 65–69.
2. Kandiyali, R.; Sarginson, J.; Hollén, L.; Spickett-Jones, F.; Young, A. The management of small area burns and unexpected illness after burn in children under five years of age—A costing study in the English healthcare setting. *Burns* **2018**, *44*, 188–194. [[CrossRef](#)] [[PubMed](#)]
3. Davé, D.R.; Nagarjan, N.; Canner, J.K.; Kushner, A.L.; Stewart, B.T.; Group, S.R. Rethinking burns for low & middle-income countries: Differing patterns of burn epidemiology, care seeking behavior, and outcomes across four countries. *Burns* **2018**, *44*, 1228–1234. [[PubMed](#)]

4. Siddharthan, T.; Grigsby, M.R.; Goodman, D.; Chowdhury, M.; Rubinstein, A.; Irazola, V.; Gutierrez, L.; Miranda, J.J.; Bernabe-Ortiz, A.; Alam, D. Association between household air pollution exposure and chronic obstructive pulmonary disease outcomes in 13 low-and middle-income country settings. *Am. J. Respir. Crit. Care Med.* **2018**, *197*, 611–620. [[CrossRef](#)] [[PubMed](#)]
5. Bailey, M.; Sagiraju, H.; Mashreky, S.; Alamgir, H. Epidemiology and outcomes of burn injuries at a tertiary burn care center in Bangladesh. *Burns* **2019**, *45*, 957–963. [[CrossRef](#)] [[PubMed](#)]
6. Mashreky, S.R.; Shawon, R.A.; Biswas, A.; Ferdoush, J.; Unjum, A.; Rahman, A.F. Changes in burn mortality in Bangladesh: Findings from Bangladesh Health and Injury Survey (BHIS) 2003 and 2016. *Burns* **2018**, *44*, 1579–1584. [[CrossRef](#)]
7. Rivas, E.; Herndon, D.N.; Chapa, M.L.; Cambiaso-Daniel, J.; Rontoyanni, V.G.; Gutierrez, I.L.; Sanchez, K.; Glover, S.; Suman, O.E. Children with severe burns display no sex differences in exercise capacity at hospital discharge or adaptation after exercise rehabilitation training. *Burns* **2018**, *44*, 1187–1194. [[CrossRef](#)]
8. Khatib, M.; Jabir, S.; Fitzgerald O'Connor, E.; Philp, B. A systematic review of the evolution of laser Doppler techniques in burn depth assessment. *Plast. Surg. Int.* **2014**, *2014*. [[CrossRef](#)]
9. Abubakar, A.; Ugail, H. Discrimination of Human Skin Burns Using Machine Learning. In *Intelligent Computing—Proceedings of the Computing Conference*; Springer: London, UK, 2019; pp. 641–647.
10. Cirillo, M.D.; Mirdell, R.; Sjöberg, F.; Pham, T.D. Time-Independent Prediction of Burn Depth using Deep Convolutional Neural Networks. *J. Burn Care Res.* **2019**, *40*, 857–863. [[CrossRef](#)]
11. Jan, S.N.; Khan, F.A.; Bashir, M.M.; Nasir, M.; Ansari, H.H.; Shami, H.B.; Nazir, U.; Hanif, A.; Sohail, M.J.B. Comparison of Laser Doppler Imaging (LDI) and clinical assessment in differentiating between superficial and deep partial thickness burn wounds. *Burns* **2018**, *44*, 405–413. [[CrossRef](#)]
12. Wearn, C.; Lee, K.C.; Hardwicke, J.; Allouni, A.; Bamford, A.; Nightingale, P.; Moiemmen, N.J.B. Prospective comparative evaluation study of Laser Doppler Imaging and thermal imaging in the assessment of burn depth. *Burns* **2018**, *44*, 124–133. [[CrossRef](#)] [[PubMed](#)]
13. Barbedo, J.G.A. Plant disease identification from individual lesions and spots using deep learning. *Biosyst. Eng.* **2019**, *180*, 96–107. [[CrossRef](#)]
14. Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; Beck, A.H. Deep learning for identifying metastatic breast cancer. *arXiv* **2016**, arXiv:1606.05718.
15. Khan, S.; Islam, N.; Jan, Z.; Din, I.U.; Rodrigues, J.J.C. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognit. Lett.* **2019**, *125*, 1–6. [[CrossRef](#)]
16. Abubakar, A.; Ugail, H.; Bukar, A.M.; Smith, K.M. Discrimination of Healthy Skin, Superficial Epidermal Burns, and Full-Thickness Burns from 2D-Colored Images Using Machine Learning. In *Data Science*; CRC Press: Boca Raton, FL, USA, 2019; pp. 201–223.
17. Abubakar, A.; Ugail, H.; Bukar, A.M. Noninvasive assessment and classification of human skin burns using images of Caucasian and African patients. *J. Electron. Imaging* **2019**, *29*, 41002. [[CrossRef](#)]
18. Wood, J.; Brown, B.; Bartley, A.; Cavaco, A.M.B.C.; Roberts, A.P.; Santon, K.; Cook, S. Reducing pressure ulcers across multiple care settings using a collaborative approach. *BMJ Open Qual.* **2019**, *8*, e000409. [[CrossRef](#)] [[PubMed](#)]
19. Santiso, L.; Tapking, C.; Lee, J.O.; Zapata-Sirvent, R.; Pittelli, C.A.; Suman, O.E. The epidemiology of burns in children in Guatemala: A single center report. *J. Burn Care Res.* **2020**, *41*, 248–253. [[CrossRef](#)]
20. Boscarelli, A.; Fiorenza, V.; Chiaro, A.; Incerti, F.; Mattioli, G.; Gandullia, P. Esophageal stricture as a complication after scald injury in children. *J. Burn Care Res.* **2020**, in press. [[CrossRef](#)]
21. Endom, E.E.; Giardino, A.P. Skin Injury: Bruises and Burns. In *A Practical Guide to the Evaluation of Child Physical Abuse and Neglect*; Springer: Cham, Switzerland, 2019; pp. 77–131.
22. Skin Bruising Images. Available online: <https://www.shutterstock.com/search/skin+bruising> (accessed on 20 February 2020).
23. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F.-F. Imagenet: A large-scale hierarchical image database. In *Proceedings of the Computer Vision and Pattern Recognition*, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; NeurIPS: Vancouver, BC, Canada, 2012; pp. 1097–1105.

25. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
27. Xiao, T.; Liu, L.; Li, K.; Qin, W.; Yu, S.; Li, Z. Comparison of transferred deep neural networks in ultrasonic breast masses discrimination. *BioMed Res. Int.* **2018**, *2018*. [[CrossRef](#)]
28. Alves, F.R.V.; Vieira, R.P.M. The Newton Fractal's Leonardo Sequence Study with the Google Colab. *Int. Electron. J. Math. Educ.* **2019**, *15*, em0575. [[CrossRef](#)]
29. Cho, H.; Matthews, G.J.; Harel, O. Confidence intervals for the area under the receiver operating characteristic curve in the presence of ignorable missing data. *Int. Stat. Rev.* **2019**, *87*, 152–177. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).