aArticle

Comparison of Deep Transfer Learning Techniques in Human Skin Burns Discrimination

Aliyu Abubakar [1,2*], Mohammed Ajuji [2] and Ibrahim Usman Yahya [2]

[1]Centre for Visual Computing, Faculty of Engineering and Informatics, University of Bradford, United Kingdom; a.abubakar6@bradford.ac.uk

[2]Department of Computer Science, Faculty of Science, Gombe State University, Nigeria

[3]Department of Computer Science, Faculty of Science, Gombe State University, Nigeria

* Correspondence: a.abubakar6@bradford.ac.uk; Tel.: +447574758316

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 pre-trained deep learning models due to deficient dataset to train a new model from scratch. 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 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 nation 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 [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. The negative situation in those countries 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 [5]. 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.

Assessment of burn injury clinically by experience personnel via visual observation is very subjective and totally dependent on the experience of the personnel with recorded high accuracy of

approximately 75% [8], in addition of been painful due to invasive nature (blanching). Other diagnostic means such as Laser Doppler Flowmetry (LDF) Laser Doppler Imaging (LDI) were found to be effective than clinical evaluation. LDF works based on principle that a laser light when shined on the skin gets reflected due to frequency shift that correlates to microvascular circulation and allowing the perfusion rate to be determined on flux scale [9]. The disadvantage of using LDF is associated to heated probe and direct contact with burn site thereby leading to painful assessment and sampling error. LDI operates similar to LDF but it is non-invasive method that gives depth analysis of the burn wound. However, the adoption of LDI is very rare; only in high income countries as a result of affordability cost and takes days to accurately produce desired outcome [10] and inaccurate for assessment in children which raises concern due to movement effects.

Moreover, the use of both LDF and LDI devices is primarily centred on diagnosing burn injuries. Certain skin injuries that may easily be misjudged by health practitioners due to close similarity includes pressure ulcer and contusion were not given consideration. Pressure ulcer injuries are mostly sustained on body locations where there is bony prominence coupled with contraction of the skin/tissue with external hard surface. These injuries are mostly developed by elderly individuals, people with physical disabilities and patients with severe illnesses whose mobility is very limited. Study has shown that over 700 000 people get affected by pressure ulcer in United Kingdom every year, costing health service more than £3.8 million per day and estimated sum of £1.4 - £2.4 billion per year [11]. Pressure ulcer can prolong hospital-stay, for instance 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.

Contusions (bruises) are mostly 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. Those contusions occurring at bony prominences such as knees, elbows, anterior tibiae and forehead are mostly due to accidental causes [12]. 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.

Towards the end, in this paper is a pipeline approach to discriminate burns and PUB injuries using deep learning the research enormous experiments and analysis using different deep learning models, and different experiments strategies. In both experimental processes, comparisons in terms of accuracy and time efficiency are well presented. The rest of the paper is organised as follows: section 2 provides a review to some related literatures; section 3 presents the 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 [8].

Study by [8] to discriminate burns and healthy skin using off-the-shelf features was proposed. The study utilised 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. 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 categorised into classes depending on whether the injury affect top-most layer, intermediate or deep layer. In this regard, authors in [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 authors in [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 VGG-16 and recorded 86.7% accuracy, 79.4% for GoogleNet, 82.4% for ResNet-50 and 88.1% for ResNet-101.

Studies by [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.

Considering all of the aforementioned studies, the idea is quite simple and was largely concentrated on discriminating whether a given abnormal image is burn or healthy. This left a very crucial and challenging gap that needs to be addressed as whether all the given abnormal images are burn images

or some sort of skin abnormalities. This is the gap we propose to address in this paper using pre-trained deep learning models.

3. Materials and Methods

The main concept here is to classify skin burn images and other skin injuries with 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 use to train a few new layers or machine learning algorithm such as support vector machine and decision tree.

3.1. Dataset

Our database contains two sets of datasets: 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. 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. We pre-processed the images, which is crucial step to ensure unwanted noise is minimised for computational efficiency improvement. Unnecessary background in the images were cropped out and normalised. Augmentation was applied to enlarge the size of the database via the use of different transformation processes such as rotation and flipping, resulting to a larger database containing 1280 images in each category.

Figure 1. Sample of dataset images

3.2. Feature Extraction

Deep learning models are greedy algorithms that require huge data to be trained effectively. Having such data pose a big challenge to earlier deep learning researchers, in addition to lack of powerful computational hardware. With the emergence of large databases such as ImageNet [18] and hardware-accelerated devices such as graphical processing unit (GPU) respectively, deep learning research was reshaped. For example, about a decade ago when the ImageNet database was made accessible by research communities, we 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 datasets to categories 1000 objects in 2012 [19], VGGNet from researchers in Oxford University and GoogleNet from Google were proposed in 2014 [20] . Going deeper in building deep learning models was challenged by vanishing gradient in which ResNet by Microsoft Research Asia in 2015 [21] proposed solution to ensure that the performance of top layers is as good as the lower layers without vanishing gradient and optimisation problem. This was achieved by using identity connection in parallel to the main convolution layers as illustrated in figure 2, making back-propagation easier and learning process faster and most importantly allows building network deeper enough to extract complex patterns from the data and results in considerable increase in accuracy. ResNet has multiple variation of architectures including ResNet50, ResNet101 and ResNet152.
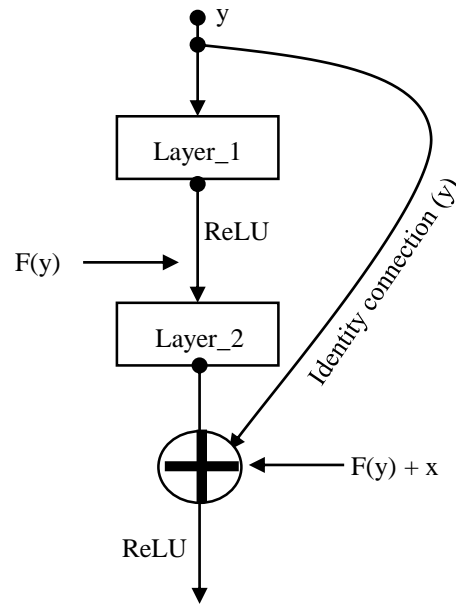
Figure 2. Illustration of ResNet identity connection

Interestingly, those big models can be re-used to achieve desired outcome in fields with deficient data via a process called transfer learning [22]. Transfer learning is to re-use learned layers of a deep learning model from one research field to another area. Deep learning model can re-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. Therefore, we use the three pre-trained ResNet models (that is, ResNet50, ResNet101 and ResNet152) in this paper using fine-tuning and off-the-shelf feature extraction approaches.

3.2.1.   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 used as illustrated in figure 3.
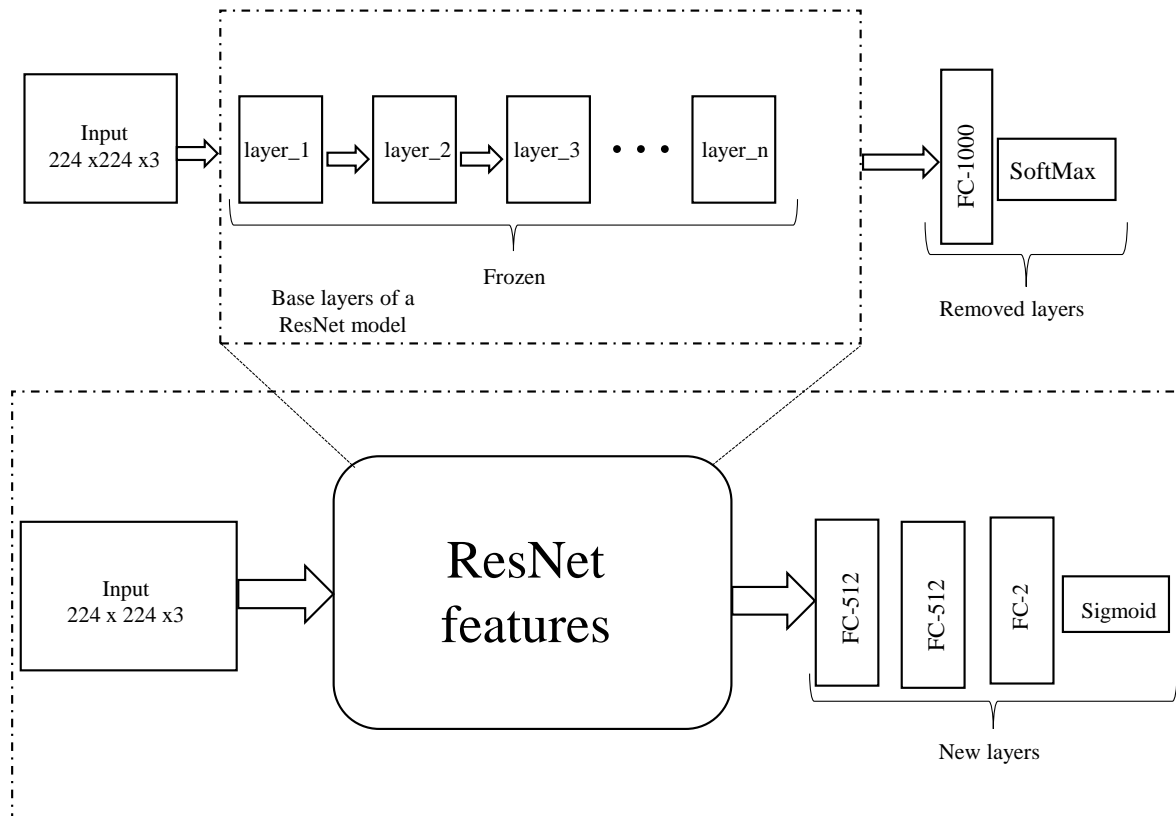
Figure 3. Illustration of fine-tuning

3.2.2. *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. Given a training set (x_i, y_i) where $i = 1, \ldots, n$ with $x \in R^N$ and $y_i \in \{+1, -1\}$, where burn images are denoted by +1 and PUB images are denoted by -1. The SVM is trained to find the most optimum separating hyperplane that discriminates the two classes of images by solving

$$\begin{cases} \min_{w,\,b} & \frac{1}{2}||w||^2 \\ \\ with & y_i.(w.x + b) \geq 1 \end{cases} \qquad (1)$$

Thus, the optimum separating hyperplane was computed which discriminated Burns and PUB into their respective category. A $K-fold$ cross validation technique was used during the classifier's training, where a value of 10 was chosen K.
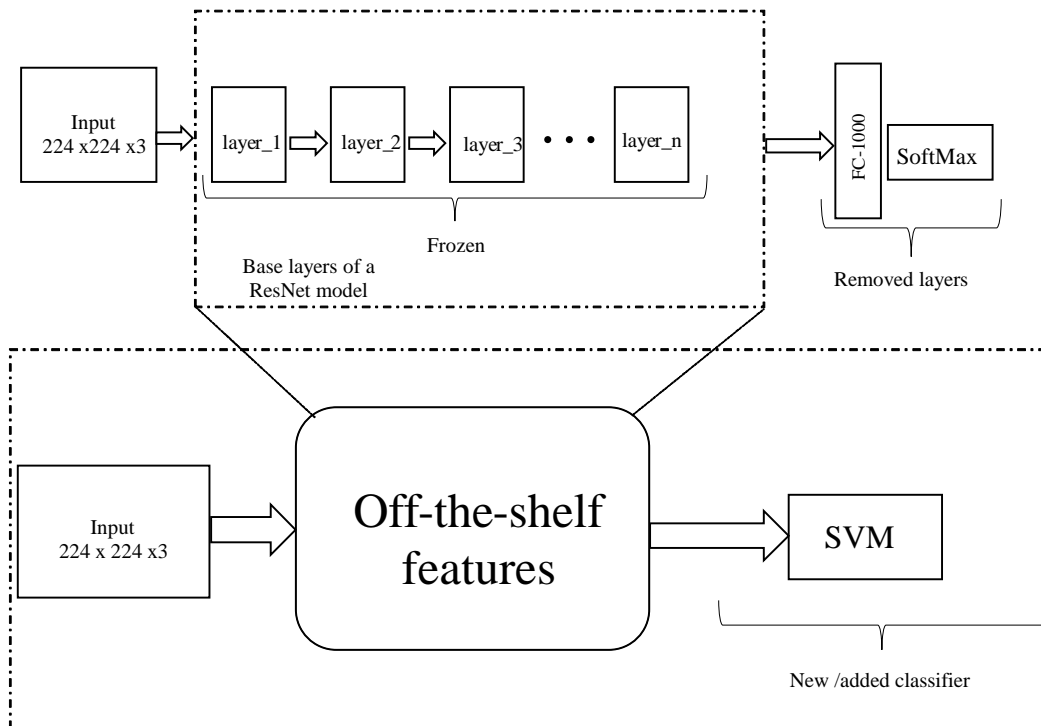
Figure 4. Illustration of off-the-shelf feature extraction and classification

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 [23]. This resource provides free access to on-demand GPU hardware-accelerator.

Figures 5, 8, and 11 depicted the training processes of fine-tuned ResNet50, ResNet101 and ResNet152 respectively, using 200 epochs in each. Figure 6 shows training and validation accuracies using ResNet50, in which the result shows a near perfect classification output, while the result depicted in figure 7 provides training and validation losses of the fine-tuned ResNet50.

4.1. Results from fine-tuning strategy

```
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
```

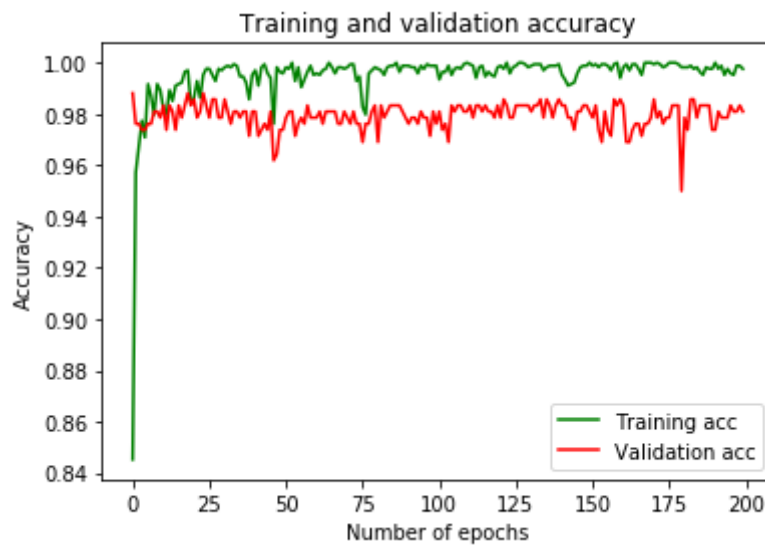**Figure 5.** Showing ResNet50 training process using fine-tuning

Figure 6. Showing training and validation accuracies using ResNet50
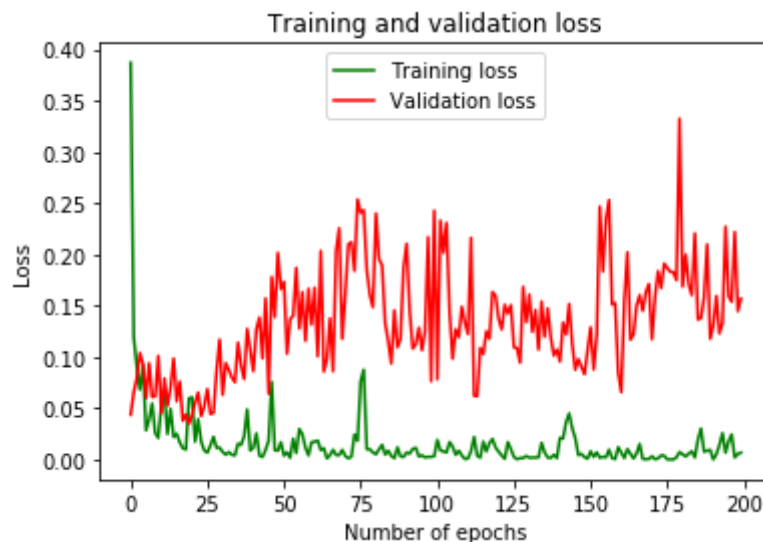

Figure 7. Showing training and validation losses using ResNet50

For ResNet101, training and validation accuracies were shown in figure 9, figure 10 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.

```
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
```

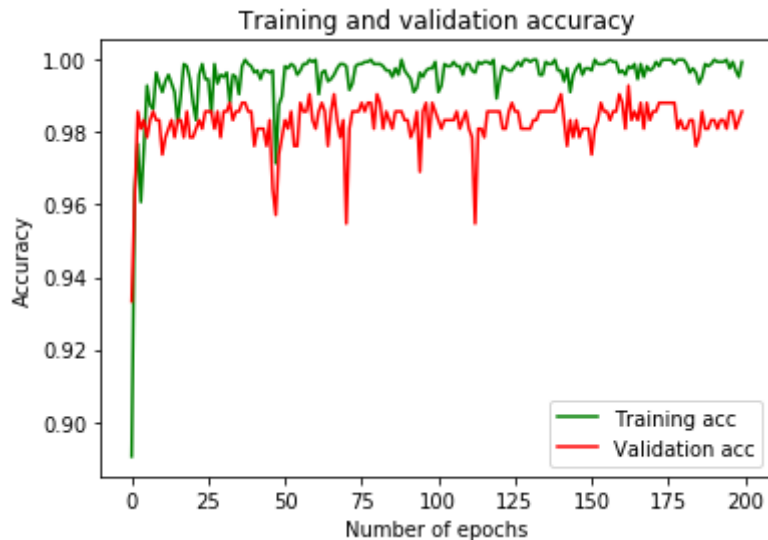**Figure 8**. Showing ResNet101 training process using fine-tuning

Figure 9. Showing training and validation accuracies using ResNet101
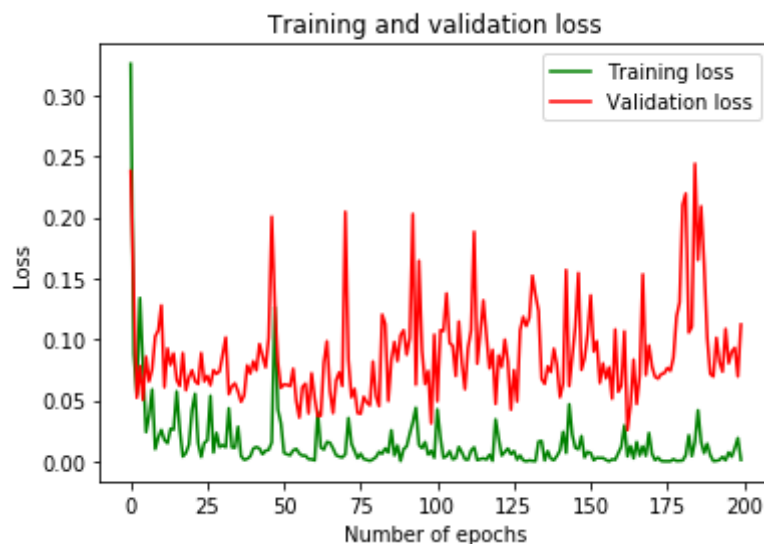

Figure 10. Showing training and validation losses using ResNet101

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.

```
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
```

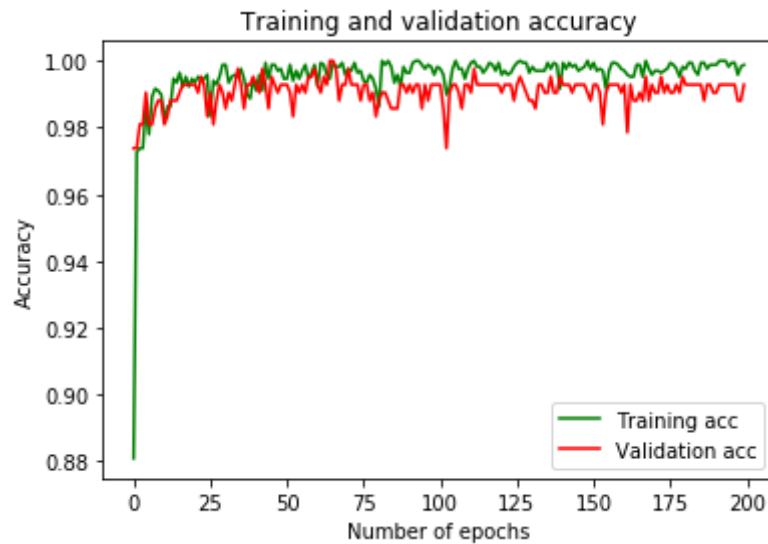**Figure 11.** Showing ResNet152 training process using fine-tuning

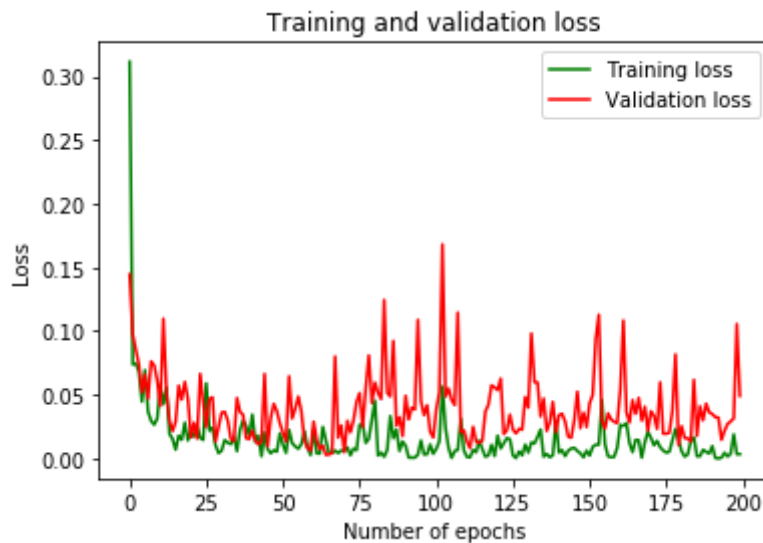**Figure 12.** . Showing training and validation accuracies using ResNet152

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. SVM classification using ResNet50 features

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

Table 2. SVM classification using ResNet101 features

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

Table 3. SVM classification using ResNet152 features

		Target Classes	
		PUB	**Burns**
Output Classes	**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 [24]. Sensitivity is determined by the following mathematical equation: $\text{sensitivity} = \frac{TP}{TP+FN} \; X \; 100$; specificity is determined by the following mathematical equation; $\text{specificity} = \frac{TN}{TN+FP} \; X \; 100.$ 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: $\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \; X \; 100.$ 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 seconds to complete compared to 2116 seconds and 2705 seconds 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(Sec.)	Accuracy	Time(Sec.)
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, Receiver Operating Characteristics (ROC) curve is another important tool for analysing diagnostic performance of a binary classifier. ROC curve gives the trade-off between sensitivity and specificity as the threshold varies [25]. 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.
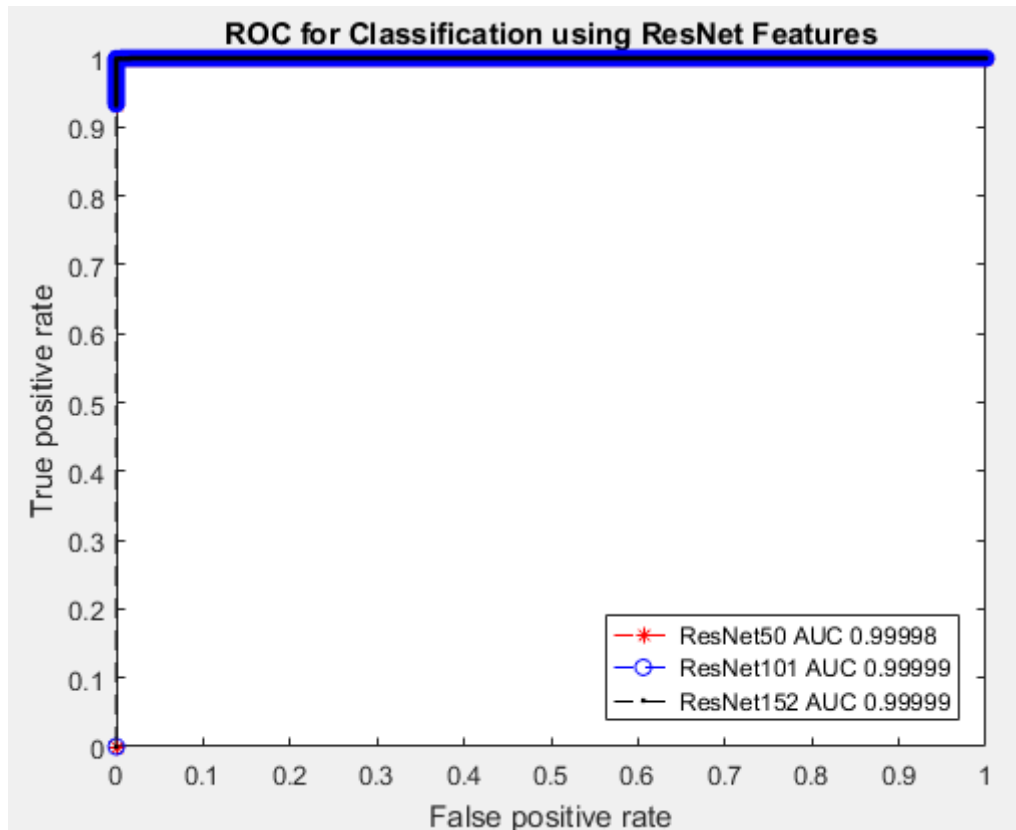
Figure 14. ROC curve for the binary classification of Burns and PUB

Moreover, though studies in the literatures 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 [8]	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, 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 seconds, ResNet101 achieved 98.8% accuracy in 2116 seconds while resnet152 achieved 99.5% accuracy in 2704 seconds. 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 models 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 seconds using ResNet152 features and 99.8% accuracy in 14 seconds and 15 seconds 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. Also, burns in human skin regardless of their type such as whether superficial, intermediate or full-thickness, they can be discriminated effectively using machine learning as the alternative where human expertise is lacking

References

[1]   G. Dexter, S. Patil, K. Singh, M. A. Marano, R. Lee, S. J. Petrone, and R. S. Chamberlain, "Clinical outcomes after burns in elderly patients over 70 years: A 17-year retrospective analysis," *Burns*, vol. 44, no. 1,   pp. 65-69, 2018.

[2]   R. Kandiyali, J. Sarginson, L. Hollén, F. Spickett-Jones, and A. Young, "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*, vol. 44, no. 1,   pp. 188-194, 2018.

[3]   D. R. Davé, N. Nagarjan, J. K. Canner, A. L. Kushner, B. T. Stewart, and S. R. Group, "Rethinking burns for low & middle-income countries: differing patterns of burn epidemiology, care seeking behavior, and outcomes across four countries," *Burns*, vol. 44, no. 5,   pp. 1228-1234, 2018.

[4]   T. Siddharthan, M. R. Grigsby, D. Goodman, M. Chowdhury, A. Rubinstein, V. Irazola, L. Gutierrez, J. J. Miranda, A. Bernabe-Ortiz, and D. Alam, "Association between household air pollution exposure and chronic obstructive pulmonary disease outcomes in 13 low-and middle-income country settings," *American journal of respiratory and critical care medicine*, vol. 197, no. 5,   pp. 611-620, 2018.

[5]   M. Bailey, H. Sagiraju, S. Mashreky, and H. Alamgir, "Epidemiology and outcomes of burn injuries at a tertiary burn care center in Bangladesh," *Burns*, vol. 45, no. 4,   pp. 957-963, 2019.

[6]   S. R. Mashreky, R. A. Shawon, A. Biswas, J. Ferdoush, A. Unjum, and A. F. Rahman, "Changes in burn mortality in Bangladesh: Findings from Bangladesh Health and Injury Survey (BHIS) 2003 and 2016," *Burns*, no., 2018.

[7]   E. Rivas, D. N. Herndon, M. L. Chapa, J. Cambiaso-Daniel, V. G. Rontoyanni, I. L. Gutierrez, K. Sanchez, S. Glover, and O. E. Suman, "Children with severe burns display no sex differences in exercise capacity at hospital discharge or adaptation after exercise rehabilitation training," *Burns*, vol. 44, no. 5,   pp. 1187-1194, 2018.

[8]   A. Abubakar and H. Ugail, "Discrimination of Human Skin Burns Using Machine Learning," in *Intelligent Computing-Proceedings of the Computing Conference*, 2019, pp. 641-647.

[9]   A. Holland, H. Martin, and D. Cass, "Laser Doppler imaging prediction of burn wound outcome in children," *Burns*, vol. 28, no. 1,   pp. 11-17, 2002.

[10]   M. D. Cirillo, R. Mirdell, F. Sjöberg, and T. D. Pham, "Time-Independent Prediction of Burn Depth using Deep Convolutional Neural Networks," *Journal of Burn Care & Research*, no., 2019.

[11]   J. Wood, B. Brown, A. Bartley, A. M. B. C. Cavaco, A. P. Roberts, K. Santon, and S. Cook, "Reducing pressure ulcers across multiple care settings using a collaborative approach," *BMJ open quality*, vol. 8, no. 3,   p. e000409, 2019.

[12]   E. E. Endom and A. P. Giardino, "Skin Injury: Bruises and Burns," in *A Practical Guide to the Evaluation of Child Physical Abuse and Neglect*, vol.: Springer, 2019, pp. 77-131.

[13]   J. G. A. Barbedo, "Plant disease identification from individual lesions and spots using deep learning," *Biosystems Engineering*, vol. 180, no.   pp. 96-107, 2019.

[14]   D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv preprint arXiv:1606.05718*, no., 2016.

[15]   S. Khan, N. Islam, Z. Jan, I. U. Din, and J. J. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognition Letters*, vol. 125, no.   pp. 1-6, 2019.

[16]   A. Abubakar, H. Ugail, A. M. Bukar, and K. M. Smith, "Discrimination of Healthy Skin, Superficial Epidermal Burns, and Full-Thickness Burns from 2D-Colored Images Using Machine Learning," in *Data Science*, vol.: CRC Press, 2019, pp. 201-223.

[17]   A. Abubakar, H. Ugail, and A. M. Bukar, "Noninvasive assessment and classification of human skin burns using images of Caucasian and African patients," *Journal of Electronic Imaging*, vol. 29, no. 4,   p. 041002, 2019.

[18]   J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248-255.

[19]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[20]   C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.

[21]   K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[22]   T. Xiao, L. Liu, K. Li, W. Qin, S. Yu, and Z. Li, "Comparison of transferred deep neural networks in ultrasonic breast masses discrimination," *BioMed research international*, vol. 2018, no., 2018.

[23]   F. R. V. Alves and R. P. M. Vieira, "The Newton Fractal's Leonardo Sequence Study with the Google Colab," *International Electronic Journal of Mathematics Education*, vol. 15, no. 2,   p. em0575, 2019.

[24]   A. Abubakar and H. Ugail, "Discrimination of Human Skin Burns Using Machine Learning," 2019, pp. 641-647.

[25]   H. Cho, G. J. Matthews, and O. Harel, "Confidence intervals for the area under the receiver operating characteristic curve in the presence of ignorable missing data," *International Statistical Review*, vol. 87, no. 1,   pp. 152-177, 2019.