Neuroscience Informatics 2 (2022) 100034

Contents lists available at ScienceDirect

Neuroscience Informatics

www.elsevier.com/locate/neuri

Artificial intelligence in brain informatics

Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer

Karar Ali [a,c,1], Zaffar Ahmed Shaikh [a,1], Abdullah Ayub Khan [a,b,1], Asif Ali Laghari [b,*,1]

[a] Faculty of Computing Science & Information Technology, Benazir Bhutto Shaheed University, Lyari, Karachi, Sindh, Pakistan
[b] Department of Computer Science, Sindh Madressatul Islam University, Karachi, Sindh, Pakistan
[c] NexDegree Private Limited, Block, House #421, 3 Siraj ud-Daulah Rd, Bahadurabad Bahadur Yar Jang CHS, Karachi, Karachi City, Sindh, Pakistan [2]

ARTICLE INFO

Article history:
Received 24 October 2021
Received in revised form 1 December 2021
Accepted 6 December 2021

Keywords:
Convolutional neural networks
CNN
Deep learning
EfficientNet
HAM10000 dataset
Medical imaging
Multiclass skin cancer classification
Skin cancer classification
Transfer learning

ABSTRACT

Skin cancer is one of the most prevalent and deadly types of cancer. Dermatologists diagnose this disease primarily visually. Multiclass skin cancer classification is challenging due to the fine-grained variability in the appearance of its various diagnostic categories. On the other hand, recent studies have demonstrated that convolutional neural networks outperform dermatologists in multiclass skin cancer classification. We developed a preprocessing image pipeline for this work. We removed hairs from the images, augmented the dataset, and resized the imageries to meet the requirements of each model. By performing transfer learning on pre-trained ImageNet weights and fine-tuning the Convolutional Neural Networks, we trained the EfficientNets B0-B7 on the HAM10000 dataset. We evaluated the performance of all EfficientNet variants on this imbalanced multiclass classification task using metrics such as *Precision*, *Recall*, *Accuracy*, *F1 Score*, and *Confusion Matrices* to determine the effect of transfer learning with fine-tuning. This article presents the classification scores for each class as *Confusion Matrices* for all eight models. Our best model, the EfficientNet B4, achieved an *F1 Score* of 87 percent and a Top-1 Accuracy of 87.91 percent. We evaluated EfficientNet classifiers using metrics that take the high-class imbalance into account. Our findings indicate that increased model complexity does not always imply improved classification performance. The best performance arose with intermediate complexity models, such as EfficientNet B4 and B5. The high classification scores resulted from many factors such as resolution scaling, data enhancement, noise removal, successful transfer learning of ImageNet weights, and fine-tuning [70–72]. Another discovery was that certain classes of skin cancer worked better at generalization than others using Confusion Matrices.

1. Introduction

According to the World Health Organization report, skin cancer is diagnosed in one out of three people worldwide. Furthermore, one in every five Americans will develop skin cancer during their lifetime, according to the Skin Cancer Foundation [1,75,76]. Melanoma and non-melanoma skin cancers are the most common types. Worldwide, approximately two to three million non-melanoma skin cancers and 132000 melanoma skin cancers are diagnosed each year [1,2].

Melanoma is the deadliest type of skin cancer. A melanoma cell tends to travel to other body parts, including the lungs, liver, spleen, or brain [2,62]. Mainly, metastatic melanoma ranks third most typical origin of central nervous system (CNS) metastases. In particular, the advanced stage of melanoma can often cause brain metastasis, whose treatment requires radiation therapy and immunotherapy. Furthermore, melanoma accounts for almost 10 percent of brain metastasis [3–5]. Melanoma is responsible for 10,000 deaths each year in the United States alone [6]. These figures appear bleak, but detecting cancer at an early stage reduces the risks of death significantly. Melanoma can be cured in nearly 95 percent of cases if detected early [7]. Thus, it is critical for early-stage diagnosis of skin cancer to prolong patient survival.

The dermatologist's experience limits the visual evaluation of dermatoscopic images (or manual dermatoscopy). Due to the subjectivity of human decision-making, besides considerable inter-class similarity in skin lesions and other confounding factors, this method is prone to mistakes. General diagnostic procedures for identifying skin cancer, such as the ABCD (Asymmetry, Border, Color, Diameter) rule [8] or the 7-point checklist [9], can only

* Corresponding author.
E-mail address: asifalilaghari@gmail.com (A.A. Laghari).
[1] These authors contributed equally to this work.
[2] https://g.co/kgs/LMViMh.

https://doi.org/10.1016/j.neuri.2021.100034

guide the practitioner with certain thumb rules. Medical professionals can misinterpret and misclassify the same dermatoscopic image sample as belonging to different kinds of skin cancer. Thus, an automated computational system needs to undertake large quantities of visual exploration using past data to imitate medical experts' expertise better and maybe outperform them too, i.e., a data-driven strategy.

With the current resurgence of interest in machine learning, deep learning, and neural networks, automated skin cancer classification has been an important topic of research [10–16]. Therefore, we leveraged the power of deep Convolutional Neural Networks (CNN) for skin lesion classification. Many machine learning-based techniques for binary classification of melanoma and non-melanoma exist. Sánchez-Reyes et al. [11] have employed HSV color space, mathematical morphology, and a Gaussian filter for ROI identification to estimate four descriptors (symmetry, edge, color, and size) in dermatological and straightforward images. Using k-Fold Cross-Validation [66,67] and a multilayer perceptron as a classifier for malignant and benign melanoma, they attained *Accuracy* of 98.5 percent and 98.6 percent, respectively. Alquran et al. [12] have segmented and extracted characteristics using the Gray Level Co-occurrence Matrix and the ABCD rule followed by Principal Component Analysis. Finally, they employed a support vector machine to classify the data. They achieved an *Accuracy* of 92.1 percent. Almaraz-Damian et al. [13] suggested a CAD system that incorporated characteristics based on texture, ABCD rule, and support vector machine for classification. Jafari et al. [14] developed an efficient pre-screening technique for binary classification of pigmented skin lesions in two groups of Melanoma and Benign. They employed directed filtering to enhance border detection of the lesion and then applied the ABCD rule for feature extraction. After that, they fed those discriminative features into a support vector machine for classification.

The preceding research has exhibited remarkable effectiveness for binary classification of Melanoma and Benign skin cancer. However, because of the significant inter-class similarity and intra-class variability, multiclass skin cancer classification (also known as automated skin cancer classification) remains problematic.

The success of neural networks in the ImageNet Large Scale Visual Recognition Challenge [10] fostered various deep learning-based solutions for multiclass skin lesion classification. Codella et al. [17] have used sparse coding, support vector machines, and deep learning to achieve 93.1 percent *Accuracy* when analyzing archived images from the International Skin Imaging Collaboration (ISIC). These imageries corresponded to *mel* (i.e., melanoma), *bkl* (i.e., Benign), and *nv* (i.e., Non-Vascular) (i.e., Atypical Nevi) (i.e., Atypical Nevi) (i.e., Atypical Nevi). Hosny et al. [18] employed data-augmentation and transfer-learning on AlexNet and achieved an *Accuracy* of 95.91 percent on the three-class ISIC multiclass dataset. Nugroho et al. [19] exploited the HAM10000 dataset to construct a skin cancer identification system utilizing custom CNN. They utilized a scaled image of 90×120 resolution. The *Accuracy* they attained was 78 percent. Bassi et al. [20] followed a deep-learning strategy on the HAM10000 dataset, i.e., transfer learning and fine-tuning. They resized HAM10000 dataset photos into 224×224 and applied a fine-tuned VGG-16 model. The *Accuracy* they attained was 82.8 percent. Moldovan et al. [21] employed a transfer-learning-based technique on the HAM10000 dataset and made an *Accuracy* of 85 percent in the first stage. Çevik et al. [22] scaled photos to 400×300 and leveraged VGGNET architecture to build a bespoke CNN. They obtained an *Accuracy* of 85.62 percent.

In this paper, we investigate the classification performance of the EfficientNets B0-B7 on the HAM10000 dataset of dermatoscopic images [23,24]. The dataset consists of 10015 images belonging to seven skin cancer classes: *akiec*, *bcc*, *bkl*, *df*, *mel*, *nv*, and *vasc*. We used ImageNet pre-trained weights to perform transfer learning and fine-tune the CNNs for the HAM10000 dataset. *Precision*, *Recall*, *Accuracy*, *F1 Score*, *Specificity*, *Roc Auc Score*, and *Confusion Matrices* evaluated the EfficientNets B0-B7 performance on this imbalanced multiclass classification task. This paper also presents the per-class classification exactitudes in the form of *Confusion Matrices* for all eight models. In particular, our best model, the EfficientNet B4, achieved an 87 percent F1 Score and 87.91 percent *Top-1 Accuracy*. Our findings indicate the superior performance of the EfficientNets B0-B7 for multiclass skin cancer classification on the HAM10000 dataset.

Table 1
Dataset distribution.

Diagnostic category	Number of images	Percentage
akiec	327	3.27%
bcc	514	5.13%
bkl	1099	10.97%
df	115	1.15%
mel	1113	11.11%
nv	6705	66.95%
vasc	142	1.42%

The remainder of the paper is structured as follows. The HAM10000 dataset and its distribution appear in the Dataset section. The Methodology section depicts the research approach. EfficientNets B0-B7 implementation and training details emerge in the Implementation section. The performance evaluation measures for fine-tuning CNNs appear in the Performance Evaluation Metrics section. The Results and Discussion section explains the study's findings. Finally, the Conclusion section puts the article to a close.

2. Dataset

This section describes the HAM10000 dataset and its distribution for training, validation, and testing.

2.1. HAM10000 dataset

The standard HAM10000 dataset has benchmarked our approach. HAM10000 stands for Human Against Machine with 10000 training images. The final dataset consists of 10015 dermatoscopic images called ISIC archive from a training set by ISIC [23]. In the HAM10000 dataset, the pigmented skin lesion classes are *akiec*, *bcc*, *bkl*, *df*, *mel*, *nv*, and *vasc*. The number and percentages of images in each class appear in Table 1. It is quite evident that there is a high-class imbalance in the dataset, with more than two-thirds of the imageries belonging to the *nv* class.

2.2. Dataset distribution

We divided the HAM10000 dataset into three parts: training (72 percent), validation (8 percent), and testing (20 percent). The Testing set helped assess the effectiveness of our trained models. We made sure no image had duplicates when it came to Validation and Testing sets. Table 2 shows three sets of the class-wise distribution of the HAM10000 dataset.

3. Methodology

This section explains the preprocessing image pipeline that we built to remove hairs from images, augment the dataset, and resize images according to the requirements of each model of the EfficientNets B0-B7. The EfficientNet model architecture, modification in model architectures, and the transfer-learning process, which trains the HAM10000 dataset on pre-trained weights of ImageNet and fine-tunes CNNs, are also explained.

Table 2
Class wise distribution of the HAM10000 dataset.

Diagnostic category	Training	Validation	Testing
akiec	48	30	49
bcc	370	48	96
bkl	775	90	234
df	82	8	25
mel	883	85	145
nv	4745	550	1410
vasc	108	12	22
Total	*7211*	*823*	*1981*

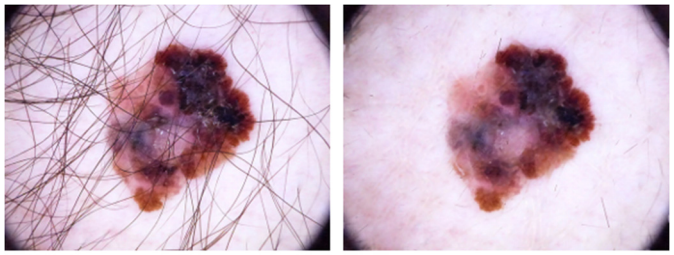

Fig. 1. Preprocessing pigmented skin lesion images: (a) original image for each class (b) preprocessed image for each class.

3.1. Image preprocessing pipeline

In the HAM10000 dataset, each image has a dimension of 600×450. The images were resized (resolution scaling) based on the EfficientNet [24,39] variant used for training.

As images in the HAM10000 dataset consist of pigmented skin lesions, our goal, i.e., the classification of skin cancer classes, the presence of hairs is not relevant. The fur in the image contributes to the noise. CNN will have to learn that the arbitrary strands spread across the skin lesion image are irrelevant to our task. Also, there is a danger for the CNN model to discover correlations between noise and the target (class of skin cancer) (class of skin cancer). If we do not remove this noise from the image, CNN will have to learn about ignoring the noise by gradient descent across a large dataset of images. Due to limitations in the size of the dataset (only 10015 images) and computation steps, image preprocessing removed most of the noise while preserving the signal in the image (Fig. 1) using image inpainting [25,72–74]. The algorithm relies on the Fast Marching Method [26,27]. We need to create a mask corresponding to the area, which must be inpainted (in our case, the hair strands in each image). The blackhat transform [28] worked as the mask. Through these two algorithms, we got a cleaner dataset of skin lesion images. The samples of the algorithm output for each class are in Fig. 1.

We increased the dataset size through image augmentation. The size of the dataset has usually been an issue in the medical domain as neural nets require a colossal amount of labeled data for training. Labeling the medical images is expensive and requires a qualified medical professional for the task. It is unlike other domains where non-experts can perform the labeling of the data. Previously, the importance of data augmentation for skin lesion analysis has been established [29]. We artificially augmented the dataset size through rotation, zooming, horizontal, and vertical flipping. This section explains the preprocessing image pipeline that we built to remove hairs from images, augment the dataset, and resize images according to the requirements of each model of the EfficientNets B0-B7. The EfficientNet model architecture, modification in model architectures, and the transfer-learning process, which trains the HAM10000 dataset on pre-trained weights of ImageNet and fine-tunes CNNs, are also explained.

3.2. EfficientNet model architecture

CNNs can be scaled to achieve better *Accuracy*. However, the scaling process was never thoroughly investigated. It entailed an iterative manual tuning process, either by arbitrarily increasing the depth or width of the CNN or by using a higher input image resolution. The EfficientNet family of architectures was developed by [24] with an intent to find a suitable method to scale CNNs to achieve better *Accuracy* (i.e., model performance) and efficiency (i.e., model parameters and FLOPS). The authors propose a compound scaling method that uses a fixed set of coefficients to uniformly scale width, depth, and resolution. That method allowed authors to produce efficient CNN architecture, which they named EfficientNet B0.

Further, they obtained EfficientNets B1-B7 by scaling the baseline network (i.e., EfficientNet B0) using the same compound model scaling method. Thus, [24] displays eight different scales' CNN architectures and their performances based on the ImageNet dataset [24]. While EfficientNet B0 CNN architecture has 5.3 million parameters and takes a 224×224 image as input, EfficientNet B7 has 66 million parameters and takes a 600×600 image as input.

Scaling the depth of the network allows CNNs to capture richer and more complex features. However, network training becomes more challenging due to the vanishing gradient problem [30]. Scaling the width of the network allows the network to capture more fine-grained features. It is also easy to train. Wide and shallow networks, on the other hand, are incapable of capturing high-level features. Finally, higher resolution images allow CNNs to capture finer-grained patterns. Larger images require more computational power and memory. In our experiments, we tested the performance of the eight EfficientNet models (EfficientNets B0-B7) on the HAM10000 dataset [23].

3.3. Transfer learning

Transfer learning, also known as domain adaptation, is a high-level concept that utilizes the knowledge acquired in a domain or task to solve related tasks. We leveraged this previously learned knowledge from the models trained on the ImageNet dataset and used their parameters for our task. However, our approach evaluates EfficientNet models on a medical image dataset of pigmented skin lesions. Due to the difference in the domains of the dataset images, we cannot directly use the pre-trained weights for inference and expect high performance. Thus, we performed a fine-tuning process. In this step, the [trained] model's parameters are tweaked precisely to adapt to the new domain of the images.

There are many ways to do fine-tuning. These include fine-tuning all or some parameters of the last few layers of a pre-trained model [31,32] or utilizing a pre-trained model as a fixed features' extractor from which features better feed each classifier, i.e., support vector machine for classification [33]. We employed both transfer learning and fine-tuning in EfficientNets B0-B7.

3.4. Modifications in network architecture

The top three layers of EfficientNet models (EfficientNets B0-B7) were suitable for the ImageNet dataset. Therefore, new layers for our use case (seven-class skin cancer prediction) replaced the top three layers. In particular, EfficientNets B0-B6 were overfitting with the top three-layer structure. For this reason, we realized the necessity to add more dense batch normalization, dropout layers at the top of each model after removing the top three layers from each model. Thus, the top three layers, i.e., Global Average Pooling 2D, dropout, and dense layers of each model, were entirely replaced with layers defined in Table 3.

Table 3
Modified layer structure for EfficientNet B0-B6.

Layer name	Layer type	Size of feature map	Activation function
Global Avg Pooling 2D	Global Average Pooling 2D	Varies for all models	N/A
Dense_1	Dense	512	swish
BatchNormalization_1	BatchNormalization	512	N/A
Dropout_1	Dropout (0.5)	512	N/A
Dense_2	Dense	256	swish
BatchNormalization_2	BatchNormalization	256	N/A
Dropout_2	Dropout (0.5)	256	N/A
Dense_output	Dense	7	softmax

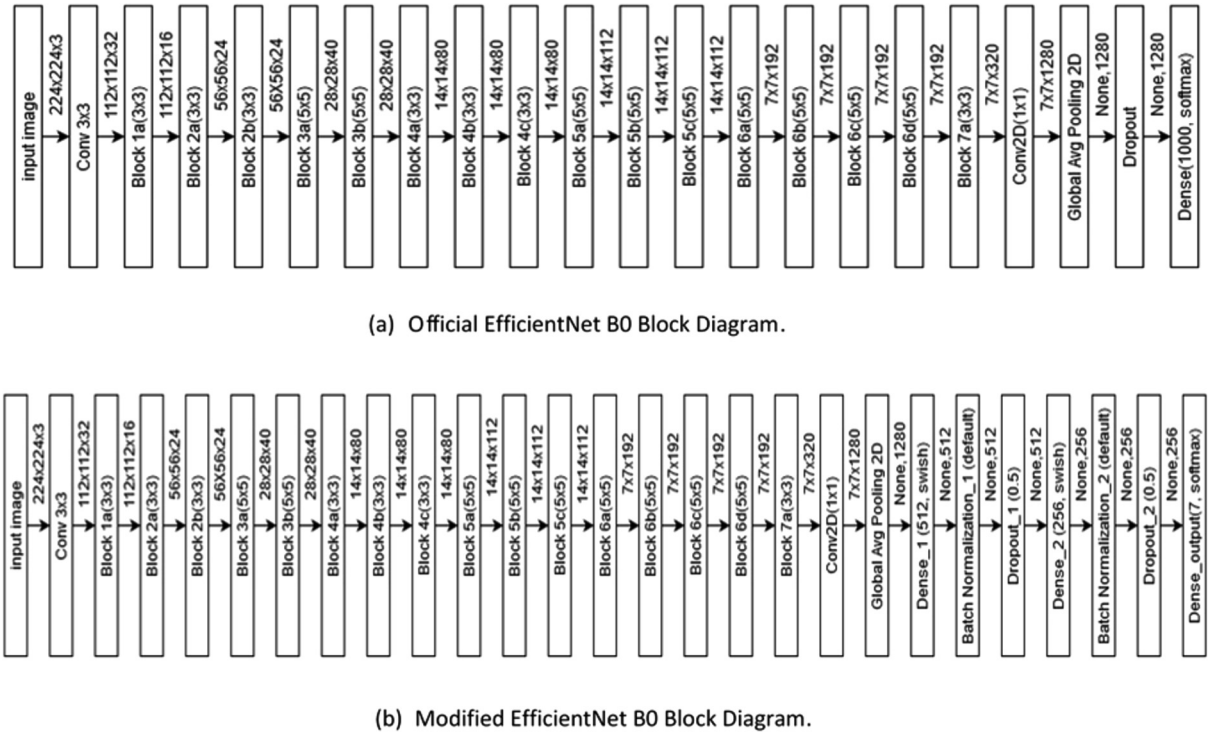

Fig. 2. The official and modified block diagrams of EfficientNet B0.

The visualized modification for EfficientNet B0 is in Fig. 2. The figure shows the block diagram of the official EfficientNet B0 baseline network and the improvements we made at the top of EfficientNet B0 architecture highlighted with a blue border. The base model (i.e., feature extractor blocks) was kept unchanged; instead, we modified the top layers of EfficientNet B0 architecture. The official B0 network has 3 top layers (i.e., global average pooling 2D, dropout, and dense layer), causing the model to overfit. We modified top layers and added the additional dense, batch normalization and drop out layers at the top of B0 base architecture where for dense (i.e., fully connected) layers, we used swish activation function [34] as used by [24] instead of RELU activation function. A blue border around the top layers in Fig. 2 highlights changes. Like EfficientNet B0, all other official EfficientNet models (i.e., B1-B7) had three layers (i.e., global average pooling2D, dropout, dense) at the top where the same modifications appeared in EfficientNet B1-B6. The same eight layers (i.e., keeping the same batch normalization, dropout rate, and feature map size of dense layers) added at the top of Efficient B0 replaced the top 3 layers of EfficientNet B1-B6. Please refer to the Supporting information section at the end of this article.

For EfficientNet B7, we removed the top three layers, i.e., global average pooling 2D, dropout, and dense output layer for 1000 classes. We replaced them with global average pooling 2D, followed by a dropout of 0.5 and the dense output layer for seven classes. The modification standard used for EfficientNets B0-B6 was not suitable for EfficientNet B7. This high complexity model was severely overfitting using the two additional dense blocks. Thus, the dropout and output layer follows the global average pool to alleviate the overfitting issue. The methodology of fine-tuning EfficientNets B0-B7 is in Section 4.

The initialization of all model parameters employed ImageNet pre-trained weights.

4. Implementation

To ensure reproducibility, we provide EfficientNets B0-B7 implementation and training details in this section.

4.1. Learning rate range

One of the most critical hyper-parameters for fast and stable neural network training is the learning rate. It decides how much of the loss gradient is to be applied to the current parameters to move them in the direction of lower loss.

We obtained the reasonable bounds of the learning rate (for a model and dataset pair) using the methods described in [35]. It was done by linearly increasing the learning rate of a network (each trained for a few epochs) in a particular range of values. In our experiment, we increased the learning rate from 10-10 to 101 and observed the change in validation loss. Fig. 3 shows the final

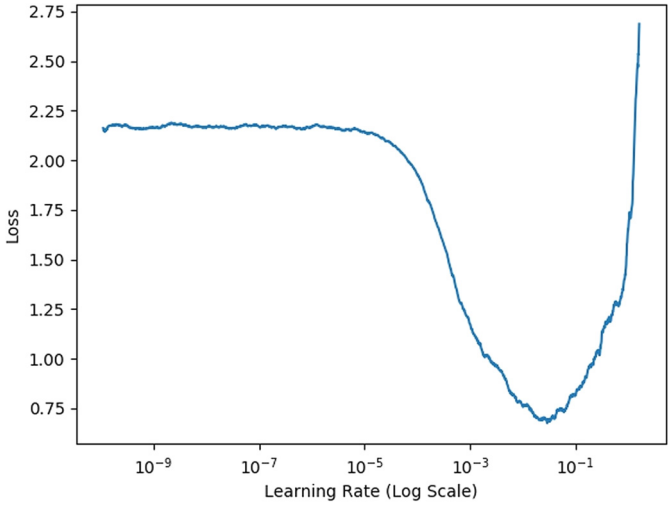

Fig. 3. Obtaining reasonable bounds of the learning rate by plotting validation loss across different learning rates.

plot of the validation loss versus the learning rate. As can be seen, the effective model training occurs between the range of values 0.0001 to 0.01, and after this range, the validation loss is high.

We used early stopping for all models [36]. The implementation relied on Keras, with TensorFlow serving as the backend. The models were trained on Google Colab, which has a 12 GB NVIDIA Tesla K80 GPU.

4.2. Fine-tuning B0-B5

For EfficientNets B0-B5, all layers (i.e., all layers received gradient updates while training) received fine-tuning, i.e., the pre-trained weights came from ImageNet, which contains images from 1000 different classes of day-to-day objects. Whereas, HAM10000 dataset is a highly domain-specific medical image dataset. Therefore, fine-tuning of all convolutional layers happened due to the significant differences in both datasets distributions. We used a Stochastic Gradient Descent (SGD) optimizer and a default learning rate decay for these models.

4.3. Fine-tuning B6-B7

For EfficientNets B6-B7, the training was more unstable in the above conditions. Thus, fine-tuning happened in two steps. In the first step, we performed fine-tuning only for newly added layers while keeping convolutional blocks frozen. It means that the convolutional base received no gradient updates. Now that the recently added layers received some weights, so for the second step, we defreezed the last four convolutional blocks of the base model while keeping all other blocks frozen and performing fine-tuning again. The last four convolutional blocks took into account computing limitations.

Also, the model complexity of the official EfficientNets B6-B7 (i.e., the model's ability to overfit on our dataset) is much larger as EfficientNets B6, and B7 contain more parameters than EfficientNets B0-B5. So, it was reasonable not to fine-tune all layers. Also, rather than using SGD, we used Adam optimizer as [37,65] did. We found that in terms of stability and performance, Adam optimizer [37,68] yielded better results than SGD while training big models like EfficientNet (B6, B7). Lastly, the polynomial decay learning rate scheduler [38,69,70] allowed for more stable convergence. We have summarized the model-specific modifications in image size, batch size, and other hyperparameters in Table 4.

5. Performance evaluation metrics

This section describes the evaluation metrics used to assess each model's performance, namely *Precision*, *Recall*, *Accuracy*, *F1 Score*, *Specificity*, *Roc Auc Score*, and *Confusion Matrices* [71].

There are no universal standards for evaluating the performance of any classification model. In literature, we see a standard set of performance measures that depend upon the user's requirement. When classes are highly imbalanced (as shown in Table 1), the metrics *Precision*, *Recall*, *Accuracy*, *F1 Score*, *Specificity*, *Roc Auc Score*, and *Confusion Matrices* are useful. As a result, we relied on them.

5.1. Confusion matrix

We created an *N*x*N* table (where *N* is the number of classes) that summarizes how well a classification model's predictions performed. It is a correlation matrix between the actual label and the classification of the model (predicted label). The results of the Confusion Matrices fall into one of four categories. A True Positive (*TP*) is when the model correctly predicts an image's positive class. When the model incorrectly predicts the positive class of an image, it generates a False Positive (*FP*). A True Negative (TN) corresponds to the case when the model correctly predicts an image's negative class. When the model mispredicts the negative class, the result is a False Negative (*FN*). In the case of a multiclass problem, the positive class is the label for which the calculation is being performed, and the negative class is the remaining label.

5.2. Accuracy

The percentage of correctly predicted image classes to the total number of images is referred to as *Accuracy* [39]. It is the most straightforward performance metric. *Accuracy* is, however, only valid when the class distribution is symmetric, that is, when the number of images (or observations) in each class is roughly the same [40]. For each of EfficientNets B0-B7, we report Top-1, Top-2, and Top-3 types of *Accuracy*. *Top-1 Accuracy* is a conventional *Accuracy*, i.e., in the model's prediction, the class with the highest probability matches the expected or the actual class [41]. *Top-k Accuracy* means that any of the model's top-k predictions based on probability must match the actual class of the image for which it is to be considered a correct prediction [42].

5.3. Precision

Precision amounts to the fraction of images correctly labeled as belonging to the positive class divided by the total number of images labeled as belonging to the positive class by the model [43,44] as follows:

$$Precision = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l} (tp_i + fp_i)}.$$

5.4. Recall

The proportion of actual positives correctly identified by the model comes from the *Recall* metric. Alternatively, it is the number of true positives divided by the number of images in the positive class [45–47]:

$$Recall = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l} (tp_i + fn_i)}.$$

Table 4
Model-specific modifications.

Model variant	Image size	Batch size*	Learning rate	Optimizer	Learning rate decay
EfficientNet B0	224×224	32	0.001	SGD	SGD decay rate
EfficientNet B1	240×240	32	0.001	SGD	SGD decay rate
EfficientNet B2	260×260	32	0.001	SGD	SGD decay rate
EfficientNet B3	300×300	16	0.001	SGD	SGD decay rate
EfficientNet B4	380×380	8	0.001	SGD	SGD decay rate
EfficientNet B5	456×456	4	0.0006	SGD	SGD decay rate
EfficientNet B6	528×528	16	0.0025	Adam	Polynomial-decay
EfficientNet B7	600×600	16	0.0025	Adam	Polynomial-decay

* The batch size was varied for higher complexity models due to limitations in computational resources.

5.5. F1 score

By the definitions of *Precision* and *Recall*, there seems to be a trade-off between the two measures. When we improve *Recall*, we reduce *Precision* and vice versa [48–50]. Depending upon the application domain and the user requirement, we might require maximizing one over the other. However, we use the *F Beta Score* in case we want the optimal blend of both metrics (i.e., to assign different weights to each metric). The *F Beta Score* amounts to the weighted harmonic mean between *Precision* and *Recall* [51]. It favors *Recall* over *Precision* by a factor of *Beta*. *Precision* and *Recall* are both equally important in this context. As a result, *Beta*=1 becomes the *F1 Score*. The harmonic mean of *Precision* and *Recall* gives the *F1 Score*. Higher *F1 Score* values indicate good predictive power [52–55]. In the case of a multiclass classification problem, we calculated the *F1 Score* over all classes to get a holistic view of the model's performance as per the expression below. Note that *F1 Score* results are not in between *Precision* and *Recall* results.

$$F1\ Score = \frac{(\beta^2+1)Precision * Recall}{\beta^2 Precision + Recall}.$$

5.6. Specificity

Specificity is a metric to determine the model's percentage of actual negative cases correctly identified as negative [56–58]. Specificity is the ratio of *TN* and the sum of *TN* and *FP*. Higher *Specificity* implies a higher *TN* value while a lower *FP* value [59–61], according to

$$Specificity = \frac{\sum_{i=1}^{l} tn_i}{\sum_{i=1}^{l}(tn_i + fp_i)}.$$

5.7. Roc_Auc score

Roc_Auc or *AUC* is also known as *AUROC*, the area under receiver operating characteristics. The *Roc_Auc Score* represents the degree of separability. It indicates how well the model can distinguish between classes. This metric usually appraises a binary classification task [62–64]. For a multiclass classification problem, one versus all methodology is used to obtain *Roc_Auc* Scores. The range of the *Roc_Auc Score* is from 0 to 1. A model having a *Roc_Auc Score* close to 1 is considered the best model. We employed one versus all methodology to obtain a weighted *Roc_Auc* Score for our models.

6. Results and discussion

Table 5 exhibits the Top-1, Top-2, and Top-3 Accuracy results of EfficientNets B0-B7 assessed on the HAM10000 dataset.

The EfficientNet B4 reports the highest *Top-1* and *Top-2 Accuracies* among all eight models. The middle-level complexity models, EfficientNets B4 and B5 beat EfficientNets B6 and B7. Finally, there are simpler models, EfficientNets B0-B3. As a result, the ranking of *Top-1 Accuracy* is B4∼B5 > B6∼B7 > B1∼B2∼B3 > B0. There is a difference of roughly 5 percent in Top-1 Accuracy performance between the best model, EfficientNet B4, and the worst model, EfficientNet B0. Across all models, the *Top-2 Accuracy* is practically the same, with a maximum absolute deviation of 1.25 percent. At the same time, the maximum absolute deviation of 0.9 percent emerges for *Top-3 Accuracy*.

Table 5
Accuracy test results of EfficientNet B0-B7.

Models	Top-1 accuracy	Top-2 accuracy	Top-3 accuracy
EfficientNet B0	83.02%	93.80%	97.39%
EfficientNet B1	83.69%	93.90%	97.34%
EfficientNet B2	83.95%	93.75%	97.39%
EfficientNet B3	83.90%	94.63%	97.65%
EfficientNet B4	87.91%	95.67%	97.81%
EfficientNet B5	87.62%	94.59%	97.55%
EfficientNet B6	85.36%	94.01%	96.97%
EfficientNet B7	85.52%	94.84%	98.12%

Table 6
Model Wise Precision, Recall, F1 Score, Specificity, and Roc_Auc Comparisons.

Models	Precision	Recall	F1 score	Specificity	Roc_Auc
EfficientNet B0	84%	83%	82%	84%	95.94%
EfficientNet B1	85%	84%	83%	84%	96.10%
EfficientNet B2	85%	84%	84%	86%	96.36%
EfficientNet B3	87%	84%	84%	91%	96.67%
EfficientNet B4	88%	88%	87%	88%	97.53%
EfficientNet B5	88%	88%	87%	88%	97.54%
EfficientNet B6	86%	85%	85%	89%	96.76%
EfficientNet B7	86%	86%	85%	87%	97.23%

As proven in the Dataset Section, the class distribution for the HAM10000 dataset is very unsymmetrical, i.e., high-class imbalance. Table 6 illustrates the *Precision*, *Recall*, *F1 Score*, *Specificity*, and *Roc Auc* scores for each EfficientNet variation on the HAM10000 dataset. Once again, the same pattern occurs for all these performance criteria, except *Specificity*.

We observed general order, i.e., B4∼B5 > B6∼B7 > B1∼B2∼B3 > B0. For *Specificity*, we noticed a different order, i.e., B3 > B6 > B4∼B5 > B7 > B2 > B0∼B1. Here, EfficientNet B3 outperformed the other models with the *Specificity* score of 91 percent. The highest performing models, i.e., EfficientNets B4 and B5, achieved a Precision of 88 percent, *Recall* of 88 percent, *F1 Score* of 87 percent, *Specificity* of 88 percent, and *Roc Auc* Score of 97.5 percent. In contrast, a dramatic difference of 5 percent has also appeared in the *F1 Score* of the best model, i.e., EfficientNet B4, and the worst model, i.e., EfficientNet B0.

The size of the dataset and model complexity justifies this observed performance trend. With models having more model complexity, there is a higher probability of attaining a better performance metric. However, at the same time, they are more prone to overfitting the dataset. HAM10000 provides little photos for a deep learning dataset (in comparison to the benchmark ImageNet

Fig. 4. Confusion matrices of EfficientNets B0-B7.

collection of 1 million images) (in contrast to the benchmark ImageNet dataset of 1 million images). Thus, outstanding performance from the witnessed mid-level complexity models (EfficientNet B4 and B5) makes intuitive sense. Lower complexity models such as EfficientNet B0-B3 have the less discriminating capability. Whereas the higher complexity models EfficientNet B6-B7 are overfitting on our dataset.

Fig. 4 illustrates the *Confusion Matrices* of EfficientNets B0-B7. Some general findings are that all models performed well on the majority class *nv* (*Accuracy* $>$ 94 percent). The photos belonging to the *akiec* class were misclassified pretty commonly as *mel* and *df* classes. Even though vasc was the minority class (only 1.42 percent of the photos belonged to this class), the average *Accuracy* across all eight models was 85.62 percent. The models, on average, fared poorly on *akiec* (34.12 percent) and *df* (24 percent) classes. They delivered best on *nv* (95.75 percent) and *vasc* (85.62 percent) classes. Thus, using *Confusion Matrices*, we notice that the perfor-

Table 7
Comparative study of the HAM10000 dataset.

Study	Preprocessing	Image type	No. of classes	Method	Accuracy
[19]	Yes	RGB	7	CNN	78%
[20]	Yes	RGB	7	CNN-transfer learning	82.8%
[21]	Yes	RGB	7	CNN-transfer learning	85%
[22]	Yes	RGB	7	CNN	85.62%
Our Proposed EfficientNet B4	Yes	RGB	7	CNN-transfer learning	87.9%

mance of eight EfficientNet models differs substantially based on the seven classifications of cancer.

In Table 7, the performance of our best model, i.e., EfficientNet B4, is compared with existing studies. They all used the same HAM10000 dataset. Nugroho et al. [19] and Bassi et al. [20] attained Accuracy of 78 percent and 82.8 percent, respectively. In comparison, Moldovan et al. [21] and Çevik et al. [22] had an *Accuracy* of 85 percent and 85.62 percent correspondingly. This comparison reveals that our suggested method has outperformed existing multiclass skin cancer classification methods.

7. Conclusion

Skin cancer is one of the most prevalent and severe cancers. This condition is primarily diagnosed visually by dermatologists. Due to the fine-grained diversity in the look of its numerous diagnostic categories, multiclass skin cancer classification is a tough undertaking [75,76]. In recent studies, on the other hand, CNNs have outperformed dermatologists in multiclass skin cancer classification. For this effort, we constructed a pretreatment picture pipeline in which we eliminated hairs from the photos, enriched the dataset, and scaled images according to each model's need. We trained the EfficientNets B0-B7 on the HAM10000 dataset by performing transfer-learning on pre-trained weights of ImageNet and fine-tuning the Convolutional Neural Networks. To analyze the influence of transfer learning and fine-tuning, we evaluated the performance of all EfficientNet variants on this imbalanced multiclass classification problem using measures such as *Precision*, *Recall*, *Accuracy*, *F1 Score*, and *Confusion Matrices*. This study shows the per-class classification scores as Confusion Matrices for all eight models. In particular, our most robust model, the EfficientNet B4, scored an 87 percent *F1 Score* and 87.91 percent *Top-1 Accuracy*.

As far as we know, this is the first study to examine EfficientNets B0-B7 performance on the HAM10000 dataset as well as the skin cancer classification challenge. We tested EfficientNet classifiers using criteria accounting for the high-class imbalance. Our results demonstrate that more model complexity does not necessarily equal better classification performance [77,78]. We noticed the best performance with middle-level complexity models such as EfficientNet B4 and B5. High classification scores resulted from a combination of reasons viz. the usage of resolution-scaling, data-augmentation, noise-removal, successful transfer-learning of ImageNet weight, and fine-tuning. Lastly, *Confusion Matrices* revealed that different classes of skin cancer showed greater generalization performance than others. It suggests space for further advancement with fine-tuned models for any given kind of malignancy.

Declaration of competing interest

The authors did not have any conflict of interest.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neuri.2021.100034.

References

[1] World Health Organization, Radiation: ultraviolet (UV) radiation and skin cancer – how common is skin cancer, https://www.who.int/news-room/q-a-detail/radiation-ultraviolet-(uv)-radiation-and-skin-cancer#, 2017. (Accessed 12 June 2021).

[2] N. Nordmann, M. Hubbard, T. Nordmann, P.W. Sperduto, H.B. Clark, M.A. Hunt, Effect of gamma knife radiosurgery and programmed cell death 1 receptor antagonists on metastatic melanoma, Cureus 9 (2017).

[3] U. Chukwueke, T. Batchelor, P. Brastianos, Management of brain metastases in patients with melanoma, J. Oncol. Pract. 12 (2016) 536–542.

[4] M.R. Lekkala, S. Mullangi, Malignant melanoma metastatic to the central nervous system, StatPearls 2020, https://www.statpearls.com/ArticleLibrary/viewarticle/24930. (Accessed 12 June 2021).

[5] S. Morais, A. Cabral, G. Santos, N. Madeira, Melanoma brain metastases presenting as delirium: a case report, Arch. Clin. Psychiatry 44 (2017) 53–54.

[6] H.W. Rogers, M.A.S. Weinstock, R. Feldman, B.M. Coldiron, Incidence estimate of non-melanoma skin cancer (keratinocyte carcinomas) in the US population 2012, JAMA Dermatol. 151 (2015) 1081–1086.

[7] M. Thörn, F. Ponté, R. Bergström, P. Sparén, H.O. Adami, Clinical and histopathologic predictors of survival in patients with malignant melanoma: a population-based study in Sweden, J. Natl. Cancer Inst. 86 (1994) 761–769.

[8] W. Stolz, A. Riemann, A.B. Cognetta, L. Pillet, W. Abmayr, D. Holzel, P. Bilek, F. Nachbar, M. Landthaler, O. Braun-Falco, ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma, Eur. J. Dermatol. 4 (1994) 521–527.

[9] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, M. Delfino, Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis, Arch. Dermatol. 134 (1998) 1536–1570.

[10] S.D. Pande, P.P. Jadhav, R. Joshi, A.D. Sawant, V. Muddebihalkar, S. Rathod, S. Das, Digitization of handwritten devanagari text using CNN transfer learning–a better customer service support, Neurosci. Inform. (2021) 100016.

[11] H. Chaves, F. Dorr, M.E. Costa, M.M. Serra, D.F. Slezak, M.F. Farez, C. Cejas, Brain volumes quantification from MRI in healthy controls: assessing correlation, agreement and robustness of a convolutional neural network-based software against FreeSurfer, CAT12 and FSL, J. Neuroradiol. 48 (3) (2021) 147–156.

[12] S.L. Bangare, Classification of optimal brain tissue using dynamic region growing and fuzzy min-max neural network in brain magnetic resonance images, Neurosci. Inform. (2021) 100019.

[13] H. Alquran, I.A. Qasmieh, A.M. Alqudah, S. Alhammouri, E. Alawneh, A. Abughazaleh, F. Hasayen, The melanoma skin cancer detection and classification using support vector machine, in: 2017 IEEE AEECT, 2017, pp. 1–5.

[14] A. Guarnizo, R. Glikstein, C. Torres, Imaging features of isolated hypoglossal nerve palsy, J. Neuroradiol. 47 (2) (2020) 136–150.

[15] K. Askaner, A. Rydelius, S. Engelholm, L. Knutsson, J. Lätt, K. Abul-Kasim, P.C. Sundgren, Differentiation between glioblastomas and brain metastases and regarding their primary site of malignancy using dynamic susceptibility contrast MRI at 3T, J. Neuroradiol. 46 (6) (2019) 367–372.

[16] J.C. Benson, V.T. Lehman, C.M. Carr, J.T. Wald, H.J. Cloft, G. Lanzino, W. Brinjikji, Beyond plaque: a pictorial review of non-atherosclerotic abnormalities of extracranial carotid arteries, J. Neuroradiol. 48 (1) (2021) 51–60.

[17] N. Codella J. Cai, M. Abedini, R. Garnavi, A. Halpern, J.R. Smith, ep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images, in: International Workshop on Machine Learning in Medical Imaging, Springer, Cham, pp. 118–126.

[18] K.M. Hosny, M.A. Kassem, M.M. Foaud, Classification of skin lesions using transfer learning and augmentation with Alex-net, PLoS ONE 14 (2019) 14.

[19] A.A. Nugroho, I. Slamet, Sugiyanto, Skins cancer identification system of HAMl0000 skin cancer dataset using convolutional neural network, Proceedings of the AIP 2019 Conference, vol. 2202, No. 1, AIP Publishing LLC, p. 020039, https://doi.org/10.1063/1.5141652.

[20] S. Bassi, A. Gomekar, Deep learning diagnosis of pigmented skin lesions, in: Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2019, pp. 1–6.

[21] D. Moldovan, Transfer Learning Based Method for Two-Step Skin Cancer Images Classification, in: 2019 E-Health and Bioengineering Conference, (EHB) 2019 Nov 21, IEEE, pp. 1–4.

[22] E. Çevik, K. Zengin, Classification of skin lesions in dermatoscopic images with deep convolution network, Avrupa Bilim ve Teknoloji Dergisi (2019) 309–318.

[23] P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data (2018).

[24] M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, arXiv preprint, arXiv:1905.11946, 2019.

[25] A. Telea, An image inpainting technique based on the fast marching method, J. Graph. Tools 9 (2004) 23–34.

[26] R.M. Sumir, S. Mishra, N. Shastry, Segmentation of brain tumor from MRI images using fast marching method, in: Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies, 2019, pp. 1–5.

[27] S.K. Siri, M.V. Latte, A novel approach to extract exact liver image boundary from abdominal CT scan using neutrosophic set and fast marching method, J. Intell. Syst. 28 (2019) 517–532.

[28] A. Yamada, A. Teramoto, K. Kudo, T. Otsuka, H. Anno, H. Fujita, Basic study on the automated detection method of skull fracture in head CT images using surface selective blackhat transform, J. Med. Imag. Health Inform. 8 (2018) 1069–1076.

[29] F. Perez, C. Vasconcelos, S. Avila, E. Valle, Data augmentation for skin lesion analysis, in: Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, in: LNCS, vol. 11041, 2018, pp. 303–311.

[30] S. Zagoruyko, N. Komodakis, Wide residual networks, in: BMVC, 2016.

[31] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the CVPR, 2014.

[32] M. Long, Y. Cao, J. Wang, M.I. Jordan, Learning transferable features with deep adaptation networks, in: Proceedings of the ICML, 2015.

[33] H. Sharif Razavian, J. Azizpour, Sullivan, S. Carlsson, CNN features off-the-shelf: an astounding baseline for recognition, in: DeepVision Workshop, Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[34] P. Ramachandran, B. Zoph, Q.V. Le, Searching for activation functions, arXiv preprint, arXiv:1710.05941, 2018.

[35] L.N. Smith, Cyclical learning rates for training neural networks, in: Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision, 2017, pp. 464–472.

[36] L. Prechelt, Early stopping-but when?, in: Neural Networks: Tricks of the Trade, 1998, pp. 55–69.

[37] A.M. Alqudah, H. Alquraan, I.A. Qasmieh, Segmented and non-segmented skin lesions classification using transfer learning and adaptive moment learning rate technique using pretrained convolutional neural network, J. Biomimet. Biomater. Biomed. Eng. 42 (2019) 67–78.

[38] Adrian Rosebrock, Deep learning for computer vision with Python, PyImageSearch, https://web.archive.org/web/20200119143500/, https://www.pyimagesearch.com/deep-learning-computer-vision-python-book/, archived on 19 January 2019.

[39] Abdullah Ayub Khan, Aftab Ahmed Shaikh, Omar Cheikhrouhou, Asif Ali Laghari, Mamoon Rashid, Muhammad Shafiq, Habib Hamam, IMG-forensics: multimedia-enabled information hiding investigation using convolutional neural network, IET Image Process. (2021).

[40] Saleemullah Memon, Kamlesh Kumar Soothar, Kamran Ali Memon, Arif Hussain Magsi, Asif Ali Laghari, Muhammad Abbas, The design of wireless portable electrocardiograph monitoring system based on ZigBee, EAI Endorsed Transact. Scalable Inf. Syst. 7 (28) (2020) e6.

[41] Asif Ali Laghari, Hui He, Muhammad Shafiq, Asiya Khan, Assessment of quality of experience (QoE) of image compression in social cloud computing, Multiagent Grid Syst. 14 (2) (2018) 125–143.

[42] Shahid Karim, Ye Zhang, Shoulin Yin, Asif Ali Laghari, Ali Anwar Brohi, Impact of compressed and down-scaled training images on vehicle detection in remote sensing imagery, Multimed. Tools Appl. 78 (22) (2019) 32565–32583.

[43] V. Shestak, D. Gura, N. Khudyakova, Z.A. Shaikh, Y. Bokov, Chatbot design issues: building intelligence with the Cartesian paradigm, Evol. Intell. (2020) 1–9.

[44] Z.A. Shaikh, Keyword detection techniques, Eng. Tech. Appl. Sci. Res. 8 (1) (2018) 2590–2594.

[45] G. Rathee, A. Sharma, H. Saini, R. Kumar, R. Iqbal, A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology, Multimed. Tools Appl. 79 (15) (2020) 9711–9733.

[46] A. Sharma, R. Kumar, Service-level agreement—energy cooperative quickest ambulance routing for critical healthcare services, Arab. J. Sci. Eng. 44 (4) (2019).

[47] Z.A. Shaikh, N. Moiseev, A. Mikhaylov, S. Yüksel, Facile synthesis of copper oxide-cobalt oxide/nitrogen-doped carbon (Cu2O-Co3O4/CN) composite for efficient water splitting, Appl. Sci. 11 (21) (2021) 9974.

[48] A. Sharma, R. Tomar, N. Chilamkurti, B.G. Kim, Blockchain based smart contracts for Internet of medical things in e-healthcare, Electron. 9 (10) (2020) 1609.

[49] Z.Y. Liu, Z. Shaikh, F. Gazizova, Using the concept of game-based learning in education, Int. J. Emerg. Technol. Learn. 15 (14) (2020) 53–64.

[50] A. Sharma, R. Kumar, Computation of the reliable and quickest data path for healthcare services by using service-level agreements and energy constraints, Arab. J. Sci. Eng. 44 (11) (2019) (Springer Science & Business Media BV).

[51] Y.W. Zhong, Y. Jiang, S. Dong, W.J. Wu, L.X. Wang, J. Zhang, M.W. Huang, Tumor radiomics signature for artificial neural network-assisted detection of neck metastasis in patient with tongue cancer, J. Neuroradiol. (2021).

[52] A. Sharma, R. Kumar, An optimal routing scheme for critical healthcare HTH services—an IoT perspective, in: ICIIP, IEEE, 2017, pp. 1–5.

[53] Z.A. Shaikh, A.A. Laghari, O. Litvishko, V. Litvishko, T. Kalmykova, A. Meynkhard, Liquid-phase deposition synthesis of ZIF-67-derived synthesis of Co3O4@ TiO2 composite for efficient electrochemical water splitting, Metals 11 (3) (2021) 420.

[54] H. Cebeci, A. Kilincer, H.İ. Duran, N. Seher, M. Şahinoğlu, H. Karabağlı, Y. Paksoy, Precise discrimination between meningiomas and schwannomas using time-to-signal intensity curves and percentage signal recoveries obtained from dynamic susceptibility perfusion imaging, J. Neuroradiol. 48 (3) (2021) 157–163.

[55] A. Sharma, G. Rathee, R. Kumar, H. Saini, V. Varadarajan, Y. Nam, N. Chilamkurti, A secure, energy- and SLA-efficient (SESE) E-healthcare framework for quickest data transmission using cyber-physical system, Sensors 19 (9) (2019) 2119.

[56] C. Kang, I.H. Lee, J.S. Park, Y. You, W. Jeong, H.J. Ahn, J.H. Min, Measuring global impairment of cerebral perfusion using dynamic susceptibility contrast perfusion-weighted imaging in out-of-hospital cardiac arrest survivors: a prospective preliminary study, J. Neuroradiol. 48 (5) (2021) 379–384.

[57] Z.A. Shaikh, S.A. Khoja, Role of teacher in personal learning environments, Dig. Educ. Rev. (2012) 23–32.

[58] A. Sharma, R. Kumar, A constrained framework for context-aware remote E-healthcare (CARE) services, Trans. Emerg. Telecommun. Technol. (2019) e3649.

[59] M. Poongodi, A. Sharma, M. Hamdi, M. Maode, N. Chilamkurti, Smart healthcare in smart cities: wireless patient monitoring system using IoT, J. Supercomput. (2021) 1–26.

[60] U.R. Kandula, D. Philip, S. Mathew, A. Subin, A.A. Godphy, N. Alex, B. Renju, Efficacy of video educational program on interception of urinary tract infection and neurological stress among teenage girls: an uncontrolled experimental study, Neurosci. Inform. 100026 (2021).

[61] Z.A. Shaikh, S.A. Khoja, Role of ICT in shaping the future of Pakistani higher education system, Turk. Online J. Educ. Technol. 10 (1) (2011) 149–161.

[62] Z. Najafpour, A. Fatemi, Z. Goudarzi, R. Goudarzi, K. Shayanfard, F. Noorizadeh, Cost-effectiveness of neuroimaging technologies in management of psychiatric and insomnia disorders: a meta-analysis and prospective cost analysis, J. Neuroradiol. (2020).

[63] A. Sharma, R. Kumar, Service level agreement and energy cooperative cyber physical system for quickest healthcare services, J. Intell. Fuzzy Syst. 36 (5) (2019) 4077–4089.

[64] Z.A. Shaikh, A.I. Umrani, A.K. Jumani, A.A. Laghari, Technology enhanced learning: a digital timeline learning system for higher educational institutes, Int. J. Comput. Sci. Netw. Secur. 19 (10) (2019) 1–5.

[65] E. Vosoughi, J.M. Lee, J.R. Miller, M. Nosrati, D.R. Minor, R.E. Abendroth, J.W. Lee, B.T. Andrews, L.Z. Leng, M. Wu, S.P. Leong, M. Kashani-Sabet, K.B. Kim, Survival and clinical outcomes of patients with melanoma brain metastasis in the era of checkpoint inhibitors and targeted therapies, BMC Cancer 18 (2018).

[66] V.V. Estrela, L.A. Rivera, P.C. Beggio, R.T. Lopes, Regularized pel-recursive motion estimation using generalized cross-validation and spatial adaptation, in: 16th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2003), 2003, pp. 331–338.

[67] T. Wong, P. Yeh, Reliable accuracy estimates from k-fold cross validation, IEEE Trans. Knowl. Data Eng. 32 (2020) 1586–1594.

[68] A.M. Coelho, V.V. Estrela, EM-based mixture models applied to video event detection, in: Principal Component Analysis - Engineering Applications, Parinya Sanguansat, IntechOpen, 2012.

[69] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philos. Trans. - Royal Soc. A, Math. Phys. Eng. Sci. 374 (2065) (2016) 20150202.

[70] A. Deshpande, P. Patavardhan, V.V. Estrela, N. Razmjooy, J.D. Hemanth, Deep learning as an alternative to super-resolution imaging in UAV systems, Ch. 9, in: V.V. Estrela, J. Hemanth, O. Saotome, G. Nikolakopoulos, R. Sabatini (Eds.), Imaging and Sensing for Unmanned Aircraft Systems, vol. 2, IET, London, UK, 2020, pp. 177–212.

[71] A. Deshpande, V.V. Estrela, P. Patavardhan, The DCT-CNN-ResNet50 architecture to classify brain tumors with super-resolution, convolutional neural network, and the ResNet50, Neurosci. Inform. (2021).

[72] S.R. Fernandes, V.V. Estrela, H.A. Magalhaes, O. Saotome, On improving sub-pixel accuracy by means of B-spline, in: Proceedings of the 2014 IEEE International Conference on Imaging Systems and Techniques (IST 2014), 2014, pp. 68–72.

[73] V.V. Estrela, N. Razmjooy, A.C.B. Monteiro, R.P. França, M.A. de Jesus, Y. Iano, A computational intelligence perspective on multimodal image registration for unmanned aerial vehicles (UAVs), in: N. Razmjooy, M. Ashourian, Z. Foroozandeh (Eds.), Metaheuristics and Optimization in Computer and Electrical Engineering, in: Lecture Notes in Electrical Engineering, vol. 696, Springer, Cham, 2021.

[74] M. Lu, S. Niu, A detection approach using LSTM-CNN for object removal caused by exemplar-based image inpainting, Electron. 9 (2020) 858.

[75] Francesco Piccialli, Vittorio Di Somma, Fabio Giampaolo, Salvatore Cuomo, Giancarlo Fortino, A survey on deep learning in medicine: why, how and when?, Inf. Fusion 66 (2021) 111–137.
[76] N. Razmjooy, M. Ashourian, M. Karimifard, V.V. Estrela, H.J. Loschi, D. do Nascimento, R.P. França, M. Vishnevski, Computer-Aided Diagnosis of Skin Cancer: A Review, Current Medical Imaging, Bentham Science Publishers, Sharjah, U.A.E, 2020.
[77] J. Hemanth, V.V. Estrela, Deep Learning for Image Processing Applications, Advances in Parallel Computing, vol. 31, IOS Press, Amsterdam, Netherlands, 2017.
[78] A. Deshpande, V.V. Estrela, N. Razmjooy, Computational Intelligence Methods for Super-Resolution in Image Processing Applications, Springer Nature, Zurich, Switzerland, 2021.