

Received July 27, 2020, accepted August 5, 2020, date of publication August 14, 2020, date of current version August 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3016651

# FCN-Based DenseNet Framework for Automated Detection and Classification of Skin Lesions in Dermoscopy Images

ADEKANMI A. ADEGUN<sup>ID</sup> AND SERESTINA VIRIRI<sup>ID</sup>, (Member, IEEE)

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban 4000, South Africa

Corresponding author: Serestina Viriri (viriris@ukzn.ac.za)

This work was supported by the University of KwaZulu-Natal, South Africa.

**ABSTRACT** Skin Lesion detection and classification are very critical in diagnosing skin malignancy. Existing Deep learning-based Computer-aided diagnosis (CAD) methods still perform poorly on challenging skin lesions with complex features such as fuzzy boundaries, artifacts presence, low contrast with the background and, limited training datasets. They also rely heavily on a suitable turning of millions of parameters which often leads to over-fitting, poor generalization, and heavy consumption of computing resources. This study proposes a new framework that performs both segmentation and classification of skin lesions for automated detection of skin cancer. The proposed framework consists of two stages: the first stage leverages on an encoder-decoder Fully Convolutional Network (FCN) to learn the complex and inhomogeneous skin lesion features with the encoder stage learning the coarse appearance and the decoder learning the lesion borders details. Our FCN is designed with the sub-networks connected through a series of skip pathways that incorporate long skip and short-cut connections unlike, the only long skip connections commonly used in the traditional FCN, for residual learning strategy and effective training. The network also integrates the Conditional Random Field (CRF) module which employs a linear combination of Gaussian kernels for its pairwise edge potentials for contour refinement and lesion boundaries localization. The second stage proposes a novel FCN-based DenseNet framework that is composed of dense blocks that are merged and connected via the concatenation strategy and transition layer. The system also employs hyper-parameters optimization techniques to reduce network complexity and improve computing efficiency. This approach encourages feature reuse and thus requires a small number of parameters and effective with limited data. The proposed model was evaluated on publicly available HAM10000 dataset of over 10000 images consisting of 7 different categories of diseases with 98% accuracy, 98.5% recall, and 99% of AUC score respectively.

**INDEX TERMS** Skin lesion, deep learning, CAD, classification, FCN, CRF, DenseNet, encoder-decoder, hyper-parameter, skin cancer.

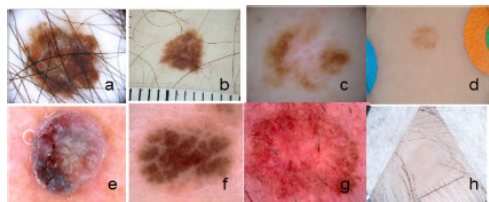
## I. INTRODUCTION

A Malignant tumor is a disorder in the human body in which unusual cells divide uncontrollably and destroy body tissue [1]. One of the prevailing malignancies in humans today is skin cancer [2] and this has been stated to be widespread in some parts of the world [3]–[6]. Among various categories of skin cancer [7]–[9], melanoma is the most deadly and dangerous form of cancer [3]. Timely identification and diagnosis of skin cancer can cure nearly 95% of cases [10]. Primarily, this disease is diagnosed visually via clinical screening and

analysis of dermoscopic, biopsy, and histopathological images [2], [10]. However, accurate diagnosis of skin lesions using these techniques is difficult, time-consuming, and error-prone even for experienced radiologists; considering the heterogeneous appearances, irregular shapes, and boundaries of the skin lesion lesions [11] as shown in Fig.1. These traditional approaches to skin lesions detection are highly intensive and laborious. They also require magnified and well-illuminated skin images for clear identification of the lesions [12], [13].

Rule-based techniques for detecting the type of skin lesions mostly employ rules such as ABCD-rule, 3-point checklist, 7-point checklist, and Menzies-rule [14], [15]. These rules

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang<sup>ID</sup>.



**FIGURE 1.** Challenging skin lesions samples: a) hair artefact, b) ruler mark artefact, c) low contrast, d) color illumination, e) bubbles, f) irregular boundaries, g) blood vessels, h) frame artefact.

have always been the foundation for diagnosis and detection by dermatologists [16], [17]. In the ABCDrule, the ABCD represents asymmetry, border structure, color variation, and diameter respectively, and asymmetry means that the two sides are unequal while symmetry means that they match. This assists in distinguishing between the benign from the malignant skin lesions. For example, the color composition is always single for Benign but can be two or more for malignant. The diameter of the general structure of the benign is always very small like a fraction of an inch but bigger and wider in malignant [17]. This dermoscopy imaging procedure is error-prone and requires years of experience in difficult situations.

Conventional methods for detecting skin lesions include thresholding methods, clustering methods, edge-based, and region-based techniques [18]. Various machine learning based-CADe systems have been designed in assisting the medicals in automated detection of skin cancer [19]. Traditional machine-learning algorithms such as gradient boosting, support vector machine (SVM) [20], artificial neural network (ANN) [21], etc have been employed by researchers for the diagnosis of skin lesions. For instance, Hameed *et al.* [22] extracted gray-level co-occurrence matrix features from skin lesions and utilized SVM to perform features classification. Murugan *et al.* [23] utilized Gaussian filters to extract lesion features and employed SVM to classify the extracted features. Seeja and Suresh [24] employed a Convolutional Neural Network (CNN) based U-net algorithm for segmentation of skin lesion and utilized a set of features extraction methods such as Local Binary Pattern (LBP), Edge Histogram (EH), Histogram of Oriented Gradients (HOG) and Gabor methods to extract color, texture, and shape features from the segmented image. The extracted features were sent into the K-Nearest Neighbor (KNN), Naïve Bayes (NB), SVM, and Random Forests (RF) classifiers to categorize them into either melanoma or benign lesions. However, since skin lesions vary in shape, size, and border features, the low-level hand-crafted methods utilized in these conventional CAD, methods possess limited discriminative capability due to their intrinsic naivety and locality. They also have other drawbacks, such as lack of adaptability in which the methods are not transferable for solving new problems [25].

In recent times, deep learning architectures have been utilized to develop computerized automated systems for detection, classification, and diagnosis of several diseases via

medical image analysis [26]. They have produced promising results most especially in the detection and classification of skin lesions cancers. They have been proven to outperform both human and existing Computer-Aided Diagnostic systems. The performance of the deep learning-based system on skin lesion detection has been evaluated against dermatologists and the conventional machine learning techniques in the recent past. Heckler *et al.* [27] explored the possibility and the advantages of using artificial intelligence for skin cancer classification against dermatologists. They established that CNN outperforms humans in the task of skin cancer classification. They employed 112 dermatologists from 13 German university hospitals and an independently well-trained CNN to classify a set of 300 biopsy-verified skin lesions into five diagnostic categories. Esteva *et al.* [11] performed classification of skin lesions using a single CNN that was trained end-to-end using only images' pixels and disease labels of skin lesions as inputs. The performance of their system was tested against 21 board-certified dermatologists on biopsy-proven clinical images. According to Brinker *et al.* [28], CNN possess the ability to classify images of skin cancer on par with dermatologists and can as well enable life-saving and quick diagnoses, through the installation of apps on mobile devices most especially outside the hospital. Guha *et al.* [29] performed experiments to compare the performance of deep learning-based techniques with traditional machine learning techniques such as SVM in the detection and classification of skin lesions. They utilized three techniques: SVM, VGGNet, and Inception-ResNet-v2, for the classification of seven categories of skin diseases.

Although existing deep learning techniques are generally more powerful than traditional methods most especially in the ability to learn highly discriminative features, their performance is still limited due to the following reasons: (1) Training deep learning methods with limited labeled data can lead to over-fitting and poor generalization. (2) Most deep learning methods require higher memory and computational resources with heavy reliant on millions of parameters tuning to perform efficiently. (3) The deep learning approach also needs to be able to process multi-scale and multi-resolution features since the skin lesion images are always acquired with different devices with varying imaging resolution. (4) Automated detection of the skin lesion is also challenging due to the heterogeneous visual attributes of skin lesions images and fine-grained contrast in the appearance of skin lesions [19].

This study proposes a new deep learning framework for automated detection and classification of skin lesion images. The proposed CAD framework consists of two main steps: the first step is detection and segmentation of skin lesions by a multi-stage encoder-decoder network and the refinement of the detected lesion border with post-processing CRF modules for better classification into various disease categories, and the second step is the classification of detected lesions with an FCN-based DenseNet system. In the first step, an encoder-decoder network was constructed to detect and segment skin

lesion of different scales and resolution, in which the encoder network connects with the decoder sub-network via series of skip pathway which is designed to integrate high-level semantic information with lower-level feature maps for efficient detection. This overcomes the problem by learning the complex and inhomogeneous skin lesion features. The system leverages on the skip pathway which is the combination of both long and short skip connections, using the short skip connections to build very deep FCNs with residual learning strategy while the long skip connections in the upsampling stage reuse the residual features to recover spatial information lost during downsampling. Specifically, in addition to the extraction of semantic features from skin lesions, our multi-stage encoder-decoder network also integrates a CRF module to further refine the extracted features for a well-defined boundary. The CRF module exploits a linear combination of Gaussian kernels for its pairwise edge potentials and efficient mean-field inference. This ensures contour refinement and lesion boundaries localization in boosting the detection performance of the classifier. In the second step, we devised an FCN-based DenseNet framework which utilized a concatenation strategy in which, the output feature maps are concatenated with the incoming feature maps to produce a large number of feature maps with a small number of convolution layers; solving the problem of a limited dataset. Also, we introduced a regularization strategy with hyperparameter optimization to train the images, which can enhance the network performance, reduce the network complexity, and improve computing efficiency for better classification performance.

The performance of the proposed model was evaluated on publicly available and standard HAM10000 dataset, which contains samplings from seven typical skin lesion categories: Melanoma (MEL), Melanocytic-Nevi (NV), Basal-Cell Carcinoma (BCC), Actinic-Keratosis and Intra-epithelial Carcinoma (AKIEC), Benign-Keratosis (BKL), Dermato-fibroma (DF), and Vascular (VASC) lesions. Standard evaluation metrics such as Accuracy, F1-Score, AUC, and Recall (Sensitivity) were used to measure the performance of the system. The results of 98% accuracy, 98.5% recall, and 99% of AUC scores respectively were obtained. Each unit of the proposed system functions independently, so we utilize the classification unit to also classify some samples of un-segmented skin lesion images from HAM10000 and PH2, and the segmented skin lesion images from ISBI 2017; the performance were compared (in the two scenarios; segmented and non-segmented lesion images).

## II. LITERATURE BACKGROUND AND RELATED WORKS

### A. BACKGROUND

Improving the performance of deep learning techniques for the analysis of skin lesion images requires a robust framework. This research examines three major factors that limit the performance of deep learning techniques in the analysis of skin lesion images: Firstly, the performance of deep learning methods is reliant on the appropriate tuning of a large number

of parameters. Most deep learning frameworks are composed of millions of parameters which directly increases the system complexity and the required computational resources [30]. Secondly, skin lesion images analysis is challenging because of the coarse visual appearances of these images which makes detection difficult [31]. These images are intricate with complex features such as fuzzy boundaries, low contrast with the background, inhomogeneous textures, or contain artifacts. Lastly, the performance of deep learning methods is primarily leveraged on large labeled datasets [32] to hierarchically learn the features that correspond to the appearance and the semantics of the skin lesion images [31]. They generally require large training data set to build efficient models and utilizing limited labeled data in a situation with skin cancer analysis can result in over-fitting and poor generalization [31]. Training deep learning methods with limited data can also lead to the generation of the coarse region of interest (ROI) detections and poor boundary definitions [30].

### B. APPROACHES AND RELATED WORKS

Lately, deep learning-based methods have been developed for the detection and classification of skin lesions into various categories of skin cancer. Various approaches and techniques of deep learning systems have been employed in the past to tackle this problem. These include methods such as transfer learning, unsupervised learning, supervised, and hybrid approaches. These approaches are however with each of them having its pros and cons:

#### 1) UNSUPERVISED LEARNING

Unsupervised fully automatic approaches have been employed in the past to tackle the problem with the scarcity of annotated medical training datasets in the analysis, segmentation, and classification of skin lesions images. Unsupervised deep learning approaches utilize strategies that derive inferences directly from a dataset which can be further used for decision making [33]. These methods generally rely on techniques such as iterative or statistical region merging, thresholding, and energy functions application [31], [34], [35]. They also utilize a probabilistic generative model with the capacity to learn the hierarchy level of features and the probability distribution over any given input space for image classification tasks [33], [36]. They, therefore, do not require large training datasets and are not in any way limited by the scarcity of annotated medical training dataset. However, they are limited in performance by the inhomogeneous appearance of medical images such as skin lesion images with the intensity distribution of the lesion containing multiple peaks. They also have a limited capacity to accurately segment challenging skin lesions, such as lesions that touch the image boundary and those with artifacts. Recently some of these methods which have been applied for medical images analysis include Restricted Boltzmann machines (RBM), Deep belief networks (DBN), Deep Boltzmann machine (DBM), Generative adversarial network (GAN) auto-encoders and its several variants [33], [34].

Semantic segmentation of medical images via the unsupervised approaches is thus challenging in producing acceptable accuracy in a life-related diagnosis. For example, Pereira *et al.* [37] developed a deep learning model that utilized the Restricted Boltzmann Machine for unsupervised feature learning of brain lesion images. They also used a Random Forest classifier for the segmentation of brain lesions. The system achieved a dice coefficient accuracy of 74% on brain MRI image datasets. Also, Akhavan Aghdam *et al.* [38] developed a deep learning algorithm using DBN for the prediction of autism. The algorithm was evaluated on Autism Brain Imaging Data Exchange I and II datasets with an average accuracy of 65.56%. They combined some series of unsupervised models comprising of rs-fMRI, GM, and WM for DBN. A Deep Neural Network (DNN) model based on Restricted Boltzmann Machine model was proposed by Al Nahid *et al.* [39] for the classification of Histopathological breast-cancer images. The system achieved an overall accuracy of 88.7% when evaluated on the breast-cancer image dataset.

Zhu *et al.* [40] presented an unsupervised classification model for MRI-based prostate cancer detection. The system achieved an averaged Section-based evaluation accuracy of 89.90% when evaluated on 21 real patient's dataset. An unsupervised model that utilized a bag of adversarial features (BAF) for the identification of mild traumatic brain injury (MTBI) in patients using their diffusion magnetic resonance images (MRI) was proposed by Minaee *et al.* [41]. The system was evaluated on a dataset of 227 samples that include 109 MTBI patients, and 118 age and sex-matched healthy controls with the mean values of over 80% accuracy on brain MRI images. Also in similar research, Vergara *et al.* [42] employed a resting-state functional network connectivity (rsFNC) model for MTBI identification. They then used a linear Support Vector Machine for the image classification. The system achieved a classification accuracy of 84.1% on extracted rsFNC features. Lastly, an experiment was performed to compare the performance of the supervised deep learning-based approach with the unsupervised deep learning-based approach on skin lesion images segmentation by Ali *et al.* [43]. It was discovered in the experiment that even though the unsupervised approach can detect the fine structures of skin lesions in some occasions, the supervised approach still produced much higher accuracy in terms of dice coefficient and Jaccard index with the supervised approach achieving a 77.7% dice coefficient score as against 40% dice coefficient score achieved by the unsupervised approach.

## 2) HYBRID LEARNING

Recently, models that employed the combination of supervised and unsupervised approaches yielded an improved performance in medical image analysis. Minaee *et al.* [44] carried out the general survey of various supervised and unsupervised methods for both semantic and instance-level segmentation. Feng *et al.* [45] for example proposed a Retinal vessel segmentation (RVS) based on a cross-connected

convolutional neural network (CcNet) for the automatic segmentation of retinal vessel trees. The system explored cross-training for model training and prediction of the pixel classes. The system was evaluated on two publicly available datasets of DRIVE and STARE with performance results of 0.7625 and 0.9528 sensitivity and accuracy scores on Drive datasets and 0.7709 and 0.9633 sensitivity and accuracy scores on the Stare dataset. A High-Resolution Network (HRNet) model which maintains high-resolution representations throughout the process of image analysis was proposed by Wang *et al.* [46] for general object detection. The proposed system has been applied in a wide range of applications, including human pose estimation, semantic segmentation, and object detection with an average 85% detection accuracy. Multi-task Framework for Skin Lesion Detection and segmentation that utilized the combination unsupervised and supervised models have been proposed by Vesal *et al.* [47]. A Multi-Class Multi-Level (MCML) classification model based on an unsupervised divide and conquer rule was developed by Hameed *et al.* [48] for medical image classification. The model explored both traditional machine learning and advanced deep learning approaches. The model was evaluated on 3672 images with a diagnostic accuracy of 96.47%. Ali *et al.* [49] proposed a model that combined the Gaussian Bayes ensemble with Convolutional Neural Network for the tasks of feature extraction and automatic detection of border irregularity from skin lesion images. The system achieved accuracy, sensitivity, specificity, and F-score results of 93.6%, 100%, 92.5%, and 96.1%, respectively when evaluated on skin lesion images dataset. Vesal *et al.* [47] proposed a faster region-based CNN (Faster-RCNN) for skin lesion images analysis. The system was composed of an unsupervised region proposal network (RPN) model for generating bounding boxes or region proposals for lesion localization in imaging. A supervised modified UNET model, SkinNet, which employed a softmax classifier was then used for the semantic segmentation of the images. The system achieved 93% for the Dice coefficient and 96% accuracy performance when trained and evaluated on ISBI 2017 and the PH2 datasets. From this literature, inferences can be made that the unsupervised approaches are still limited in medical image analysis most especially in the analysis of skin lesion images. They require millions of parameters for their architectures and thereby requiring a large number of computational resources.

## 3) TRANSFER LEARNING

Transfer learning approaches have been utilized in training supervised deep learning models for medical image analysis. This has been employed to overcome the challenges with limited training labeled dataset. Transfer learning approaches are generally effective but are suboptimal on medical images analysis due to the large discrepancy that exists with the target data in this context. This can be seen from the visual appearance of images and class labels, which may cause the feature extraction process to be biased to the source data and

eventually generalize less well on the target data [50]. This is because the models are originally pre-trained on images that are different from medical images. Some of these images may include images such as animals, automobiles, equipment, etc. which have different forms from medical images that usually possess characteristics such as fuzzy boundaries, fine-grained variability, and heterogeneous appearance. Systems based on this approach are also heavy-weight and require millions of parameters and a large number of computational resources. These challenges have limited the performance of these models on medical image analysis.

The performance evaluation of these models on medical images shows that they are still yet to outperform the state-of-the-art. For example, El-Khatib *et al.* [51] applied the transfer learning approach on CNN models which were already pre-trained on ImageNet and Places365 datasets. They also used other pre-trained models such as GoogleNet, ResNet-101, and NasNet-Large. These models were then fine-tuned on skin lesions datasets via the transfer learning approach for skin lesion images detection. The models were integrated and evaluated on skin lesion images with the accuracy scores of 88.33% 88.24% 88.46% 86.79% for Accuracy, Specificity, Sensitivity, and Dice coefficient respectively. Also, an intelligent diagnosis scheme was proposed for multi-class skin lesion classification by Hammed *et al.* [52] using a hybrid approach of deep convolution neural network and SVM based error-correcting output codes (ECOC). A pre-trained CNN model, AlexNET, was utilized for feature extraction. The system achieved an overall accuracy of 86.21% when evaluated on skin lesion image datasets. Another CNN model pre-trained on Imagenet was utilized by Almaraz-Damian *et al.* [53] for the extraction and segmentation of both handcraft and deep learning features. The system achieved similar results of 87% accuracy with the models developed by El-Khatib *et al.*

Kalouche *et al.* [54] utilized three different models: logistic regression, a deep neural network, and a pre-trained CNN VGG-16 model for skin lesion images classification. The system achieved a 78% classification accuracy on skin lesion images containing melanoma cancer. A segmentation recommender based on transfer learning and crowdsourcing algorithm was proposed by Soudani and Walid [55]. The system utilized two pre-trained CNN models based on VGG16 and ResNet50 for features extraction and classification of skin lesion images. The system achieved 78.6% accuracy with the two models when evaluated on ISIC 2017 skin lesion dataset. An automatic skin lesions classification system that employed the transfer learning approach was presented by Hosny *et al.* [56]. The proposed system was based on a pre-trained CNN model based on Alex-net architecture. The architecture weight was then fine-tuned on the ISIC skin lesion dataset. The system achieved 95.91% accuracy when evaluated on ISIC 2017 skin lesion dataset. Akram *et al.* [57] proposed another classification system based on three pre-trained CNN models: DenseNet 201, Inception-ResNet-v2, and Inception-V3. These models were integrated

and fused with an entropy-controlled neighborhood component analysis (ECNCA) algorithm for feature selection and classification of skin lesion images. The system also achieved 95.9% when evaluated on ISBI 2017 skin lesion dataset. Ahmad *et al.* [58] performed discriminative analysis and classification of features from skin disease images using the CNN model based on two pre-trained models: ResNet152 and InceptionResNet-V2. They achieved an average accuracy of 84.91% and 87.42% on ResNet152 and InceptionResNet-V2 respectively. An integrated diagnostic system that utilized segmentation techniques for optimization to improve the classification performance of deep learning models for skin lesion classification was proposed by Al-Masini *et al.* [59]. The system was based on four pre-trained CNN architectures: Inception-v3, ResNet-50, Inception-ResNet-v2, and DenseNet-201. These were integrated and evaluated on both ISIC 2016 and ISIC 2017 skin lesion datasets. The system achieved the prediction accuracies of 77.04% on ISIC 2016, and 81.29% on ISIC 2017 dataset. Finally, an in-depth analysis of several deep learning-based techniques such as a fully convolution neural network, pre-trained model, ensemble, and handcrafted methods for skin lesion analysis and melanoma detection was carried out by Naeem *et al.* [60]. They concluded that by performing fine-tuning of hyper-parameters, overfitting can be reduced and the performance of a deep learning system can be improved greatly for the analysis and diagnosis of skin lesion images.

#### 4) SUPERVISED LEARNING

Lastly, we also reviewed the supervised learning approaches that have been utilized for skin lesion analysis and detection; Esteva *et al.* [11] devised a deep learning-based method using CNN for automated classification and detection of skin lesions. They utilized a CNN model that was trained in an end-to-end approach from images' pixels and disease labels serving as inputs to achieve the classification of skin lesions. They performed two binary classifications with keratinocyte-carcinomas versus benign seborrheic-keratosis, and malignant melanomas versus benign nevi. Gessert *et al.* [61] utilized a multi-resolution ensemble of CNNs comprising of EfficientNets, SENet, and ResNeXt WSL for the detection of skin lesions. They achieved satisfactory performance on a much smaller dataset of HAM 10000 and ISIC 2018. Khalid *et al.* [56] also performed an automatic skin lesions classification system using the approach of transfer learning and the pre-trained deep neural network. The transfer learning was applied on Alex-net and the architecture's weight was fine-tuned. The system was able to detect and classify segmented color image lesions into either melanoma and nevus or melanoma, seborrheic keratosis, and nevus. Three popular skin lesion datasets; MED-NODE, Derm-IS, and Derm-Quest and ISIC were utilized for both training and testing. They obtained classification-accuracy of 96.86%, 97.70%, and 95.91% on the datasets respectively.

A segmentation methodology, FRCN, was developed for the segmentation of skin lesions by first learning the full resolution features of individual image' pixel of the input skin lesion images. The system was assessed on two publicly accessible datasets; ISBI 2017 and PH2 datasets. The proposed system attained a segmentation accuracy of 95.62% for some representative of clinical benign cases, 90.78% of melanoma cases, and 91.29% of seborrheic-keratosis cases in the ISBI 2017 dataset [62]. Ratul *et al.* [16] devised a deep learning model with dilated convolution based on transfer learning from four standard architectures; VGG16, VGG19, MobileNet, and Inception-V3. They utilized the HAM10000 dataset that comprises a total of 10015 dermoscopic images of seven skin lesion categories with large class imbalances for training, validating, and testing. They achieved a classification accuracy of 87.42%, 85.02%, 88.22%, and 89.81%, with VGG16, VGG19, MobileNet, and InceptionV3 respectively.

Shimizu *et al.* [63] proposed a method that is suitable for both melanocytic skin lesions (MSLs) and non-melanocytic skin lesions (NoMSLs). They devised a method to identify Melanomas, Nevi, BCCs, and Seborrheic-keratosis using features such as color, sub-region, and texture. They utilized both layered model and flat models to function as baselines for evaluating performance. Their method was tested on 964 dermoscopy images: 105 melanomas, 692 nevi, 69 BCCs, and 98 SKs with the layered model outperforming the flat models and achieved an accuracy of 90.48%, 82.51%, 82.61%, and 80.61% for melanomas, nevi, BCCs, and SKs, respectively. Alqudah *et al.* [64] employed both GoogleNet and AlexNet with transfer learning and optimization gradient descent adaptive momentum learning rate (ADAM) for the classification of skin lesion images. The methods were applied on the ISBI 2018 database to perform classification of images into three main categories; benign, melanoma, seborrheic keratosis under two schemes: classification of segmented and non-segmented lesion images. The overall classification accuracy of 92.2% was obtained for the segmented dataset and 89.8% was obtained for the non-segmented dataset.

Preprocessing steps such as lesion image enhancement, filtering, and segmentation were utilized on lesion images to acquire the Region-of-Interest (ROI) by Almaraz-Damian *et al.* [53]. Both handcraft features and deep learning features were extracted. ABCD rule was used to extract features such as shape, color, and texture while CNN was used to further extract the deep learning features. The CNN architecture used was first pre-trained on ImageNet. MI measurement metrics were used as fusion rules for collecting vital details from both the handcraft and deep learning features. Kawahara *et al.* [65] utilized a linear classifier that was trained on extracted features from CNN. The CNN was pre-trained on natural images to differentiate between ten skin lesions. The approach also utilized a fully convolutional network for the extraction of multi-scale features via the pooling-over of augmented feature space. The proposed approach achieved an accuracy of 85.8% over

a 5-class dataset of 1300 images. Finally, a deeply supervised multi-scale network [66] was utilized for the detection and segmentation of skin cancer from skin lesion images. They utilized the side output layers of the architecture to accumulate information from both shallow and deep layers to design a multi-scale connection block that can process various changes in cancer size. Generally, the supervised approaches perform better than the other approaches in the analysis of skin lesion images.

### C. OUR CONTRIBUTIONS

In this research, we devised a fully automatic system for skin lesion detection and classification on an FCN-based densenet framework. We propose FCN for the system optimization to achieve the following: a) to reduce the computational cost and weight size by integrating compressed convolutional blocks (via the encoder-decoder and the skip pathway approach) that are light-weight into the densenet framework; b) the encoder-decoder and the skip pathway of the FCN will also allow the system to efficiently extract skin lesion features even with the limited training dataset.

The main components of the framework that serve as our contribution include the following:

#### 1) ENCODER-DECODER SEGMENTATION APPROACH

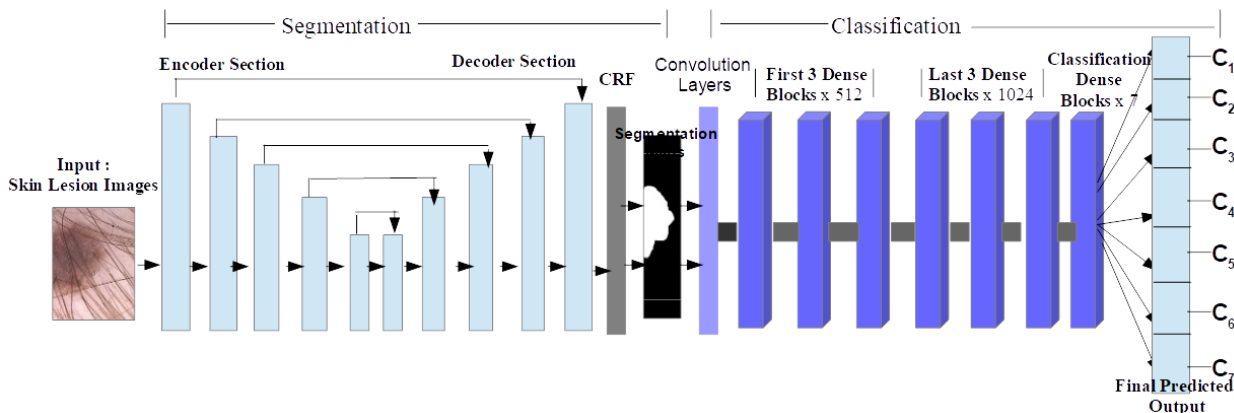
We proposed an efficient pre-processing and segmentation of skin lesions for effective features extraction by utilizing an encoder-decoder network in which the sub-networks can learn and extract the complex features of the skin lesion with the encoder stage learning the coarse appearance and localization information while the decoder learns the region based global features of the lesion. The encoder provides low-resolution features mapping and the decoder restores the features into full-resolution and further improves the boundaries delineation. This mechanism also achieves better detection and extraction of multi-scale lesion features in a limited dataset.

#### 2) RESIDUAL LEARNING STRATEGY WITH SKIP PATHWAYS

The skip pathways introduce both long skip and short-cut connections unlike the only long skip connections commonly used in the standard FCN. The system leverages the short skip connections to build very deep FCNs and as a residual learning strategy for extracting features. The long skip connections in the up-sampling stage reuse the features to recover spatial information lost during downsampling. The skip pathways can hierarchically merge both the down-sampling features with the up-sampling features and bring the semantic level of the encoder feature maps closer to that of the decoder to reliably detect lesions with flexible sizes and scales.

#### 3) INTEGRATION WITH CRF

We employed parallel integration of dense CRFs and fast mean-field inference which exploits the linear combination of Gaussian kernels for its pairwise edge potentials. This is to



**FIGURE 2.** Schematic layout diagram of our proposed deep learning framework for skin lesion images segmentation and classification where C1...C7 represents the 7-class (AKIEC, BCC, BKL, DF, MEL, NV, VASC).

ensure contour refinement and lesion boundaries localization to boost classification performance.

4) DENSENET FRAMEWORK

To develop an efficient classification system that eschews the learning of redundant feature maps to improve the classification accuracy, we devised a novel FCN-based DenseNets framework. DenseNet needs fewer parameters than a counterpart conventional CNN since it does not require learning redundant feature maps. Our proposed framework can produce selective features in a data-driven approach that can efficiently process the fine-grained unevenness in the appearance of skin lesions with a reduced computation cost.

5) CONCATENATION STRATEGY WITH TRANSITION LAYER

The output feature maps are concatenated with the incoming feature maps to produce a large number of feature maps with very little convolution. This enables us to use fewer parameters to produce a large number of feature maps thereby, overcoming the limitation with heavy reliance on a large number of parameters and datasets. The transition layers utilize a  $1 \times 1$  convolution layer between the two contiguous dense blocks for easy information transfer.

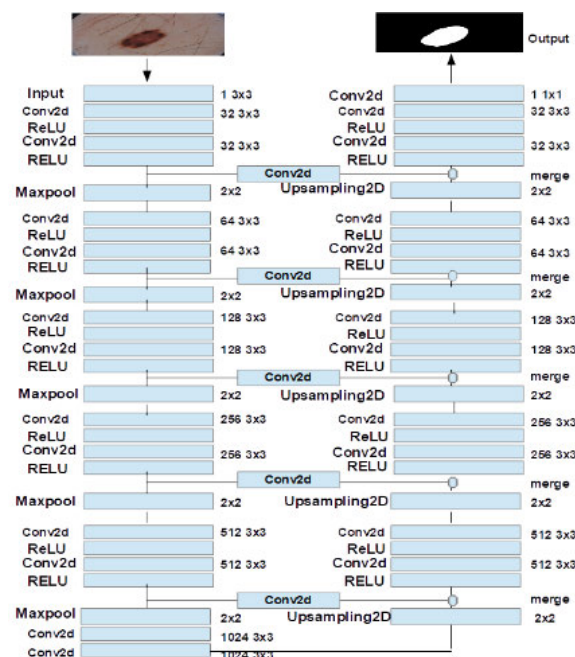
6) REGULARIZER STRATEGY AND HYPER-PARAMETERS OPTIMIZATION

The proposed system employs a regularization strategy and utilizes dropout modules in between the dense blocks. The system also performs an experimental tuning of the Hyper-parameters to enhance the network performance, reduce the network complexity, and improve computing efficiency.

III. METHODS

The methodology consists of two main components; the first component is an encoder-decoder network integrated with a fully connected CRF for lesion contour and boundaries refinement to produce highly accurate, soft segmentation maps; the second component is an FCN-based Densenet framework composing of six consecutive dense blocks with a fixed feature maps size connected with a transition layer for effective classification process. The methodology framework

of this research is described and illustrated in Figure 2 and Figure 4, and discussed within the components stated below:



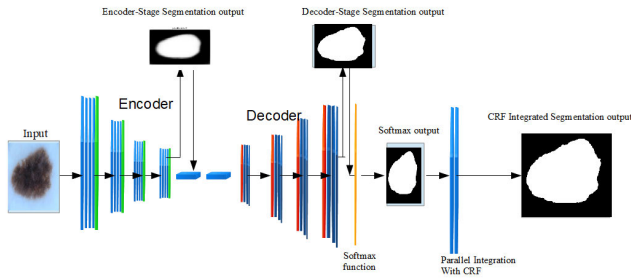
**FIGURE 3.** Architectural Diagram for Deep Convolutional Encoder-Decoder Network.

A. MULTI-SCALE FEATURE LEARNING, DETECTION AND EXTRACTION

An enhanced encoder-decoder network which is deeply supervised is employed for the task of feature learning, detection and extraction of multi-scale and multi-size skin lesion features. The composition of this network is described below:

1) FEATURE EXTRACTION WITH ENCODER-DECODER NETWORK

The network is made up of encoder and decoder sections [67] with each of the sections composed of five consecutive stages



**FIGURE 4. The Proposed framework and flow diagram of the Deep Convolutional Encoder-Decoder Network Integrated with CRFs.**

as illustrated in Figure 3. Each of the stages is made up of a convolution layer with a kernel size of 3 x 3, a ReLU activation layer, a series of skip pathways, and a concatenation layer. The number of convolutional filters increases from 64 in the first stage to 1024 in the last stage. We have replaced the usual short skip connection with a series of skip pathways which is made up of both long and short skip connections. The ReLU activation module is utilized to introduce nonlinearity which results in faster training for the network. The encoder section, in addition, utilizes max-pooling modules for down-sampling tasks. Features vectors are extracted via the convolution layers from the input images, these are then down-sampled by half using the max-pooling modules and the pooling indexes are passed to the corresponding upsampling layer in the decoder section. This is illustrated in equation 1.

$$Y_i = U(F(I : r) : d) \tag{1}$$

where  $Y_i$  is the final output,  $F$  is the downsampled feature map,  $r$  is the RELU activation function,  $d$  is the downsampling module and  $U$  is the upsampling module

The decoder section then utilizes up-sampling layer to upsample the feature vectors from the previous layers with a multiplier factor of 2. These are then concatenated with the corresponding output feature maps of the matched encoder section to achieve enriched information, avoid vanishing gradient and restore the lost feature information. The last part of the decoder section is made up of a convolutional layer with 1 x 1 kernel and softmax module to perform mapping of each pixel to a particular category of skin lesion. The softmax classifier then predicts the class for each pixel with the output in an N-channel image of probabilities and the predicted segmentation corresponded to the class with the maximum probability of each pixel. This is illustrated in equation 2.

$$P(y = i|x) = \frac{e^{x^T w_i}}{\sum_{n=1} e^{x^T w_n}} \tag{2}$$

where  $x$  is the feature map,  $w$  is the kernel operator and  $n$  represents the number of classes.

The encoder section achieves a low-resolution feature vectors and can also learn the coarse appearance and the localization details of the skin lesion while the decoder achieves

restored full-resolution feature vectors and can also learn the lesion boundaries' features. The system is also able to process efficiently multi-scale skin lesion images using a scalable framework that is adaptable and easy to modify.

2) SERIES OF SKIP PATHWAYS FOR CONNECTION

From the diagram in Figure 3, the skip pathway utilizes both long skip and short-cut connection and the system leverages on the short skip connections to build very deep FCNs and also as a residual learning strategy for efficient features extraction. The short-cut connections are made up of 2 x 2 convolution layers and they facilitate features extraction and learning. The system utilizes the series of skip pathway to hierarchically merge both the down-sampling features with the up-sampling features and bring the semantic level of the encoder feature maps closer to that of the decoder in order to reliably detect lesions with flexible sizes and scales. The long skip connections in the up-sampling stage reuse the extracted features to recover spatial information lost during downsampling.

B. FULLY CONNECTED CRF FOR POST PROCESSING

Fully connected dense CRFs with an efficient mean-field approximation and probabilistic inference are integrated into the Encoder-Decoder networks. The final output of the encoder-decoder network is then sent into the CRF module for refinement and enhancement of lesion contour, to produce the final predicted feature map and mask.

1) GAUSSIAN KERNEL EXPLORATION FOR PAIRWISE EDGE POTENTIALS IN FULLY CONNECTED CRF

The input image  $X: x1:::xN$  and the corresponding labelling mask  $Y: y1:::yN$  are taken into the CRF model in an end-to-end trainable fashion. CRF utilizes Gibbs distribution [68], a probabilistic inference model to model  $P(y|x)$  for prediction as follows in equation 3.

$$P(y|x) = \frac{1}{Z(x)} \exp[-E(y, x)] \tag{3}$$

where  $X : x1 \dots xN$  are the input features,  $Y : y1 \dots yN$ , as label mask,  $E(x|y)$  is the cost of assigning label to pixel also known as energy and  $Z$  is the constant known as partition function.

The CRF presents a probabilistic graphical model where each node represents a pixel in an image,  $I$ , and each edge represents relation between pixels. These then produce the unary and pairwise terms [69]. The unary term measures the cost of assigning label  $y$  to pixel  $x$  and it represents per-pixel classifications while the pairwise terms shows relationship between neighbouring pixels and it presents a set of smoothness constraints. The energy function is represented by  $E(x)$  represents the parameters used by unary and pairwise networks as illustrated in equation 4. The Unary potential encodes local information about a given pixel with the likelihood of a pixels to belong to a certain class such as foreground or background. The pairwise potential encode the



neighbourhood information between two neighbouring pixels and ensures smooth edges and annotations. Unary potential functions on nodes while the pairwise potentials function on edges. Assigning the most probable label to each pixel will give lower energy which implies lower cost, and thus, higher accuracy.

$$E(x) = \sum_i \psi(x_i) + \sum_{i<j} \psi(x_i, x_j) \quad (4)$$

The values of  $i$  and  $j$  in the above formula range from 1 to  $N$ , where  $X : x_1 \dots x_N$  represents input image,  $Y : y_1 \dots y_N$  represents labelling mask,  $\Psi(x_i)$  represents the unary potentials and  $\Psi(x_i, x_j)$  represents the pair-wise potentials.

### Introducing Gaussian Kernel:

We utilized Gaussian kernel function for the mean field update of all variables in the fully connected CRF model [69]. This enables the CRF model to optimize the probability map via the exploitation of local similarity of the neighbourhood pixels. The individual pixels in the unary potentials of the probability map are propagated according to their neighbourhood pixels via the pairwise potentials. A Gaussian kernel is applied to finally smoothen the boundary and to further improve the appearance kernel and smoothness kernel. The Gaussian kernel function is represented in equation 5 as:

$$k^{(m)}(f_i, f_j) = \exp\left(-\frac{1}{2}(f_i - f_j)^T \Lambda^m (f_i - f_j)\right) \quad (5)$$

where  $k^{(m)}(f_i, f_j)$  is the Gaussian Kernel function where vectors  $f_i$  and  $f_j$  are feature vectors for pixels  $i$  and  $j$  in an arbitrary feature space  $\Lambda^m$  is a symmetric positive-definite precision matrix. The pairwise potential is defined as a linear combination of Gaussian kernel in arbitrary feature space. The pairwise potentials in the model is represented in equation 6 as:

$$\psi(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^K w^{(m)} k^{(m)}(f_i, f_j) \quad (6)$$

A multi-class image segmentation with color vectors  $I_i$  and  $I_j$  is represented in equation 7 as:

$$k(f_i, f_j) = w^{(1)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta^2} - \frac{|I_i - I_j|^2}{2\theta^2}\right) + m \quad (7)$$

$$m = w^{(2)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta^2}\right) \quad (8)$$

### 2) EFFICIENT INFERENCE- MEAN FIELD APPROXIMATION

For an efficient inference in fully connected CRFs, the CRF distribution is approximated by the mean field [69]. Approximate inference program which is based on mean-field approximation is applied to minimise variational free energy. This computes a distribution  $Q(x)$  instead of the exact distribution  $P(x)$  i.e Distribution  $Q(x)$  minimises KL-divergence  $D(Q||P)$  and is expressed in equation 9 as:

$$Q(x) = N_i Q_i(X_i) \quad (9)$$

This represents the products of independent marginals  $Q_i$  and  $X_i$  respectively and  $i$  ranges from 1 to  $N$ . Performing sequential updates of  $Q_i$  will guarantee converge. The model proposes an approach to guarantee convergence with any shape of the pairwise potentials and with parallel updating using convolution mean fields together with Gaussian potentials derived from the unary and pairwise potentials.

### C. CLASSIFICATION BASED ON DenseNet SCHEME

A novel FCN based Dense-Net framework is utilized for the classification task of skin lesions into 7 categories. The structure of this framework is described below:

#### 1) DENSE BLOCKS

An efficient classification system is developed by utilizing some combination of dense blocks. These dense blocks exploit DenseNets CNN architecture which does not require learning of redundant feature maps unlike the traditional CNN that learns from redundant feature maps. The input images are first sent into 2 convolution layers with 128 and 256 output channels respectively to boost feature extraction and learning process before being sent into the dense blocks. The convolution layers have the kernel size of  $2 \times 2$  and each side of the inputs is zero-padded by one pixel to keep the feature map constant and reduce the network parameters size. The architecture is composed of six Dense Blocks with an equal number of layers; all layers with the same feature-map sizes and are connected directly with one another. The first three layers possess an output channel of 512 each and the remaining three blocks have an output channel of 1024 each. This is to ensure the utmost information flow between layers. Each of the dense blocks also consists of a dense layer, a ReLu activation function and a flatten layer to downsample the feature maps. In the model, the dense connections within the dense blocks employ sum operation for the feature merging inside the dense block to reduce the computing cost of the dense blocks. The dense blocks constructed can produce selective features in a data-driven manner to solve the problem of fine-grained variability in the appearance of skin lesions. The generated feature maps are finally processed by a 7 channel dense layer to classify the merged feature map into 7 categories of skin lesion using the sigmoid classifier. The DenseNets framework is illustrated in Figure 5.

#### 2) CONCATENATION STRATEGY

The concatenation strategy is employed to reduce extremely the number of network parameters in our proposed architecture. The layers in the dense blocks are connected to each other in a feed-forward pattern and the input feature map for each layer is concatenated with the feature maps of the preceding layers. In order to reduce the computing cost, the features for all the inner layers of the dense blocks are merged by sum operation illustrated in equation 11 while the feature maps form the input and output layer only are concatenated. The concatenation operation generates an increased number

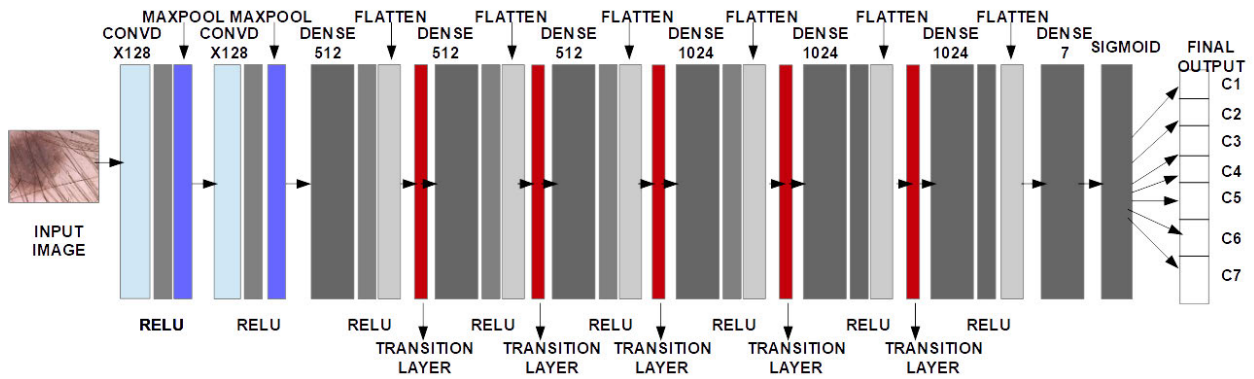


FIGURE 5. The framework diagram for the dense blocks where C1...C7 represents the 7-classes (AKIEC, BCC, BKL, DF, MEL, NV, VASC).

of feature maps with very little convolution layers which are then repeatedly used. This also reduces the number of parameters employed as it eliminates the amount of redundant feature maps learned through encouraging feature reuse. The last (nth) layer obtains the feature vectors of all previous layers  $x_0 \dots x_{n-1}$  as illustrated in equation 10.

$$x_i = k(x_0, \dots, x_{i-1}) \quad (10)$$

where  $x_i$  represents the sum operation for the feature merging within the dense block and  $k$  is the merging function.

$$y = C_n([x_0, x_1, \dots, x_{n-1}]) \quad (11)$$

where  $x_0 \dots x_{n-1}$  denotes the concatenation of the input feature-maps with the concatenation function  $C_n$

### 3) TRANSITION LAYER AND HYPERPARAMETER OPTIMIZATION

A training strategy was devised in the framework that exploits both transition layer procedure and hyper-parameter optimization technique. The transition layer is composed of a convolution layer with a kernel size of  $1 \times 1$ , a ReLU activation function, and a dropout module. This is utilized in between two neighboring dense blocks for smooth features transition. Convolutional operation is exploited to prevent vanishing gradient i.e protecting feature information from vanishing and also make the parameters of the whole framework effectively learnable. The dropout module performs a stochastic transformation on the input dimensions to avoid over-fitting.

Hyper-parameter optimization is introduced to fine-tune network parameters to optimize the system performance. The aim is to train the model faster, reduce overfitting, and make better predictions with the model. Three major optimization algorithms which include Adaptive Moment Optimization (Adam), Stochastic Gradient Descent algorithm (SGD) and Root Mean Square Propagation (RMSprop) were explored and deployed. Major hyper-parameters in the network such as learning rate, decay constant and the number of dense layers were also varied and tuned. The network was finally optimized using Adam optimizer algorithm with the following

parameters set as: (Adam optimizer = 0.0001, batch size = 128, weight decay = 0.001, drop out rate = 0.5). Experiment results are presented in Table 6 which shows the impact of varying these hyper-parameters on the system performance.

## IV. EXPERIMENTS AND RESULTS

In this section, various experiments were performed to evaluate the performance of each of the stages of our proposed framework. The segmentation stage was first evaluated, the classification stage was then evaluated and the whole system was finally evaluated. Publicly available skin lesion datasets were utilized to demonstrate the performance of each section of the system and the whole system entirely. The performance was evaluated and compared with the existing state-of-the-arts.

### A. DATASETS

The datasets used in this work can be categorized into training, validation and testing datasets:

Our training data contains 10030 images and 1 ground truth response CSV file was taken from HAM10000 (“Human Against Machine with 10000 training images”) [70] dataset. It is made up of dermatoscopic images collected from different populations under different procedures. It is a composition of important skin lesion diagnostic categories: Actinic keratoses and intraepithelial carcinoma(akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (vasc). The system was validated on the validation data that also contains 10030 images from HAM10000 dataset and also containing skin lesion diagnostic categories: Actinic keratoses and intraepithelial carcinoma(akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (vasc). The test data are taken from both ISBI 2018 [71] and PH2 [72]. The PH2 dataset is made up of 8-bit RGB color images with a resolution of  $768 \times 560$  pixels containing a total of 200 dermoscopic images of melanocytic lesions. These include 80 common nevi, 80 atypical nevi, and 40 melanomas. For the

segmentation section, we utilized ISBI 2017 [73] dataset which is composed of 2000 images and ground truth labels respectively for the segmentation model training.

### 1) DATA AUGMENTATION

We performed on the fly data augmentation on 2000 images for segmentation and the 10030 images for classification in the training dataset for both segmentation and classification process. This process was performed by applying settings such as flipping, rotation, scaling, and shear on the dataset as stated below:

- 1) Rescaling=1./255,
- 2) Shear range=0.2,
- 3) Zoom range=0.2,
- 4) Horizontal flip=True,
- 5) Rotation = random

### B. PERFORMANCE EVALUATION METRICS

The following standard metrics have been employed in this research to measure the performance of the proposed system at different stages. They are defined as stated below:

**Dice Similarity Coefficient:** It is the measures of similarity between the ground truth and predicted outcomes.

$$DSC = \frac{2TP}{FP + 2TP + FN} \quad (12)$$

**Recall (Sensitivity):** This is the proportion of actual positives that are identified correctly.

$$Sensitivity = \frac{TP}{TP + FN} \quad (13)$$

**Precision:** This is the proportion of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

**F1 Score:** This is the weighted average of Precision and Recall.

$$F1Score = \frac{2 * (Recall * Precision)}{Recall + Precision} \quad (15)$$

**Specificity:** This is the proportion of actual negatives that are identified correctly.

$$Specificity = \frac{TN}{TN + FP} \quad (16)$$

**Accuracy:** This is the proportion of correctly predicted observation(both true positives and true negatives) to the total observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$

**ROC curve:** An ROC curve (receiver operating characteristic curve) is a graph that shows the performance of a classification model. It is curve plot of Recall vs False Positive Rate.

**AUC(Area Under the ROC Curve):** It represents the complete two-dimensional area within the entire ROC curve from origin (0,0) to point (1,1).

Where  $FP$  is the amount of false-positive outcome,  $FN$  is the amount of false-negative outcome,  $TP$  is the amount of true-positive outcome and  $TN$  is the amount of true-negative outcome.

### C. RESULTS AND DISCUSSION

In this section, both the automated segmentation and classification performance of our proposed frameworks were evaluated and the results compared with the performance of the state-of-the-art methods. The performance of the segmentation unit was conducted in two phases: In the first phase, the performance of the multi-scale detection encoder-decoder network was only evaluated. This performance was evaluated against existing methods as shown in Table 1 and Table 2. In the second phase, the encoder-decoder network was integrated with the CRF modules after which the performance was again evaluated and the result was compared with the performance of the encoder-decoder network only. The segmented image outputs were compared in Figure 7 and performance metrics results were compared through the chart in Figure 6. The output was then sent into the classification network for further processing. The performance of the classification unit was also evaluated in two phases: In the first phase, the performance of the classification system was evaluated separately on un-segmented images. The results were compared with the state-of-the-art classification methods as shown in Table 3. Also, Figure 8 and Figure 9 show the accuracy performance curve and dice-coefficient curve of the system respectively. The classification performance of the

**TABLE 1. Performance evaluation(%) of the proposed model as against existing methods for segmentation process on ISBI 2017 Datest.**

Techniques	Accuracy	Dice Score	Sensitivity	Specificity
<b>Proposed model</b>	<b>95.50</b>	<b>92.1</b>	<b>97.5</b>	<b>96.5</b>
FrCN [62]	94.03	87.08	85.40	96.69
CNN-HRFB [74]	93.8	86.2	87.0	96.4
FCN [74]	92.7	82.3	84.6	96.7
iFCN [75]	95.30	88.64	85.44	98.08
DCL-PSI [35]	94.08	85.66	86.20	96.7

**TABLE 2. Performance evaluation(%) of the proposed model as against some recent semantic segmentation models for medical image analysis.**

Techniques	Jaccard Index (JI)
<b>Proposed model</b>	<b>90.50</b>
Multi-Class Multi-Level (MCML) [48]	82.0
Multi-task Framework [47]	88.0
CC-Net [76]	86.8
EXFUSE [77]	88.0

**TABLE 3. Performance Evaluation of the Proposed model Compared with Existing Methods on HAM10000 DataSet.**

Methods	ACC %	Precision %	Recall %	F1-Score
<b>Proposed model</b>	<b>98.30</b>	<b>98</b>	<b>98.5</b>	<b>98.0</b>
DCN TransferL [79]	94.92	80.36	79.8	80.36
DilatedVGG16 [16]	90.10	87	87	87
MultiLevelDN [80]	90.0	88	76	82
DilatInceptV3 [16]	90.95	89	89	89

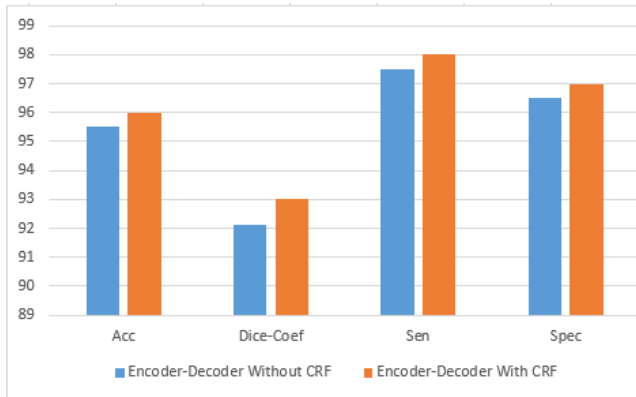


FIGURE 6. The figure compares the general performance of the encoder-decoder network using Dice-coefficient, accuracy, sensitivity and specificity when used with CRF and when used without CRF for segmentation of skin lesion images.

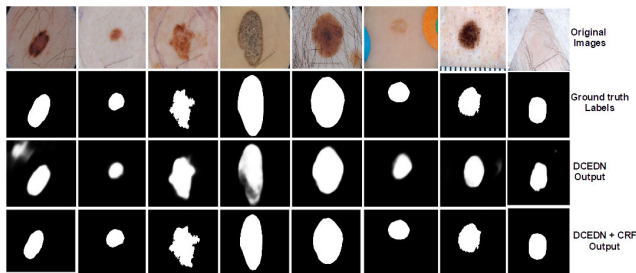


FIGURE 7. The figure shows the segmentation outcome of the encoder-decoder network when used with CRF and when used without CRF on ISBI 2017 dataset: The first row shows the Input images; the second rows shows the Ground truth labels; the third row shows segmented outcome on encoder-decoder network only; the last row shows the segmented outcome of encoder-decoder network + CRF.

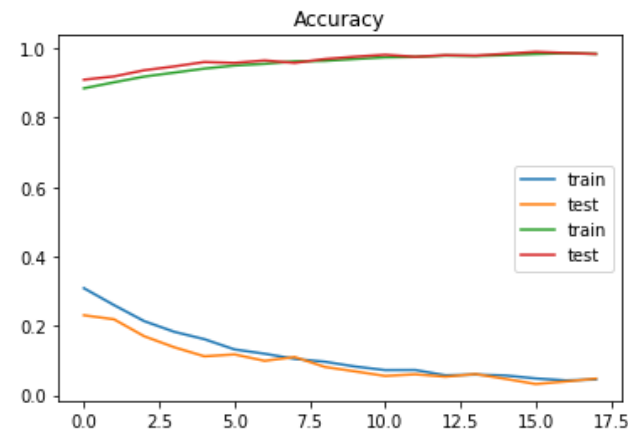


FIGURE 8. Accuracy and loss performance curves of the proposed classification model for both validation and training on HAM10000.

system on the 7 categories of skin lesion are presented with confusion matrix, ROC curve, image classification output and classification reports in Figure10, Figure 11, Figure 12, Table 4, Table 5, and Figure 13 respectively. In the second phase, the classification system was evaluated on the

TABLE 4. Classification Performance Reports (%) of the proposed model on HAM10000 Dataset.

Class Name	Precision %	Recall %	F1-Score %
AKIEC	95	98	96
BCC	99	96	98
BKL	93	100	96
DF	100	97	98
MEL	98	94	96
NV	99	99	99
VASC	100	99	100

TABLE 5. Confusion Matrix of the proposed model on HAM10000 DataSet.

Class	AKIEC	BCC	BKL	DF	MEL	NV	VASC
AKIEC	319	2	3	0	3	0	0
BCC	12	495	7	0	0	0	0
BKL	0	0	1095	0	1	3	0
DF	2	0	2	111	0	0	0
MEL	1	0	18	0	1048	45	0
NV	2	1	53	0	12	6654	0
VASC	0	0	0	0	0	1	141

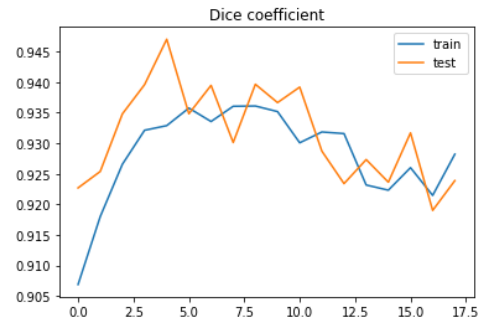


FIGURE 9. Dice-coefficient performance curves of the proposed classification model for both validation and training on HAM10000.

segmented skin lesion images from the CRF-based encoder-decoder network. The classification output of the segmented images is shown in Figure 14. Finally, the performance of the classification system with segmented skin lesion images was compared with its performance with un-segmented images as shown in the chart in Figure 16. The system was also evaluated on sample images from PH2 Dataset to test the generalization ability of the system as shown in Figure 15. Basically, the evaluation approach adopted focused on evaluation of each stage of the system as stated below:

1) SEGMENTATION AND DETECTION RESULTS

The segmentation model was trained and evaluated on augmented ISBI 2017-challenge dataset containing 2000 images and 600 images for both training and testing tasks respectively. In the first section of the experiment, we evaluate the performance of our multi-scale detection encoder- decoder network and compare its performance with state-of-the-arts among which are FrCN, CDNN, FCN, and mFCNPI methods. The evaluation was carried out on the ISBI 2017 dataset using metrics such as segmentation accuracy, dice-coefficient, sensitivity, and specificity respectively,

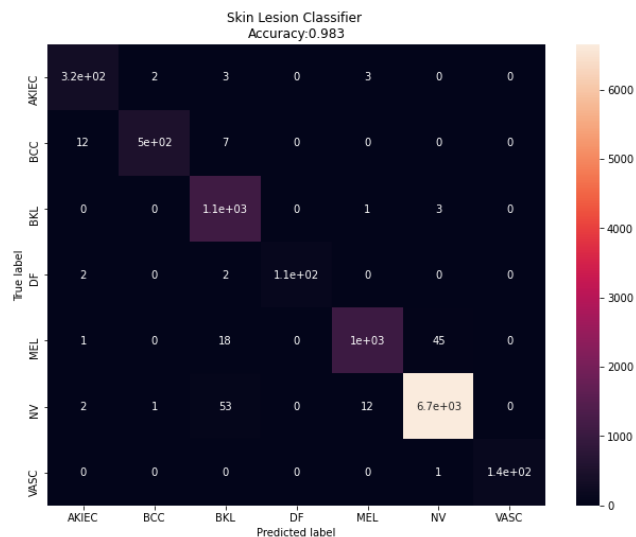


FIGURE 10. The figure shows the Confusion matrix evaluation of the proposed classification model on HAM10000.

and the corresponding results are summarized in Table 1. As shown in Table 1, we achieved the Accuracy of 95.5%, Dice Coefficient of 92.1%, Sensitivity of 97.5%, and Specificity of 96.5% on the ISBI 2017 dataset. This outperformed

Some extension of Receiver operating characteristic to multi-class

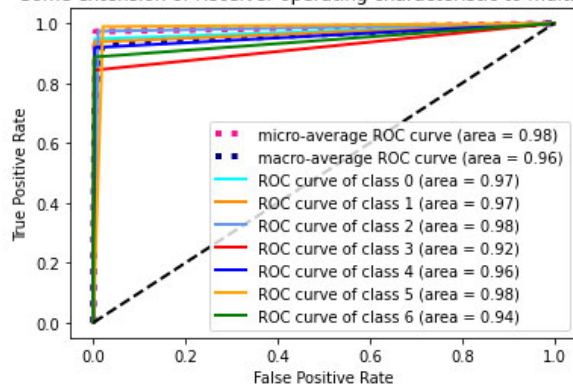


FIGURE 11. ROC performance curves of the proposed classification model for the seven categories of skin lesion in HAM10000.

some existing methods in Table 1. The performance of the proposed model was also evaluated against some recent semantic segmentation models for medical image analysis such as CC-Net, ExFuse, and Multi-class multi-level classification algorithms as shown in Table 2. The result shows the highest recall(sensitivity) and dice score of 97.5% and 92.1%

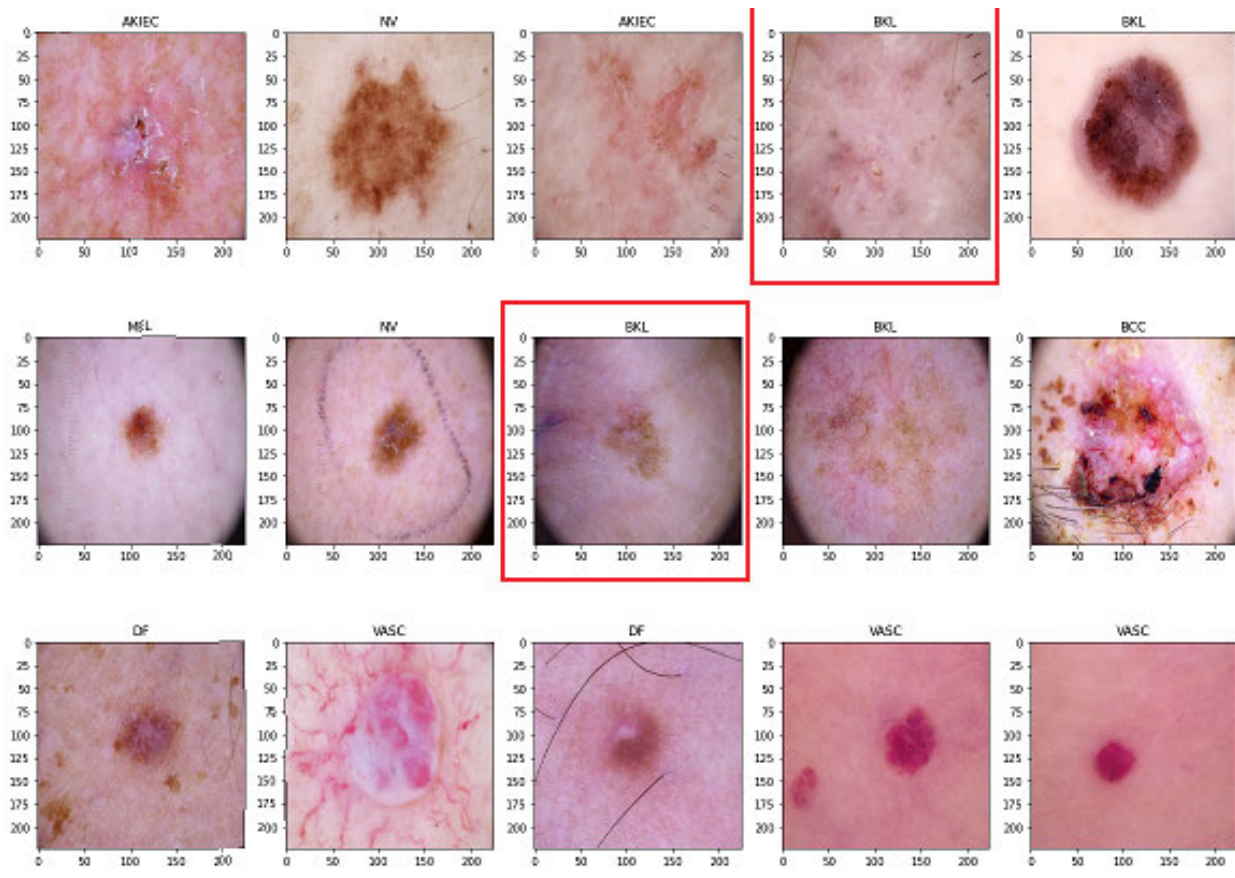


FIGURE 12. The figure shows the classification results of the proposed classification method on some sample unsegmented images from HAM10000.

when compared with other techniques. This result shows that the proposed segmentation system can detect and differentiate correctly diseased lesions from the healthy tissues on ISBI 2017 dataset as shown in Figure 6.

The encoder-decoder network integrated with the CRF model yields better performance compared with the encoder-decoder network only from the chart in Figure 6 with the overall accuracy was improved by 0.5 (96 vs. 95.5), dice-coefficient was improved by 0.9 (93 vs. 92.1), sensitivity by 0.5(98 vs. 97.5) and specificity by 0.5(97 vs. 96.5) respectively when tested on ISBI 2017. Figure 7 shows the segmentation outputs of both encoder-decoder network with CRF and without CRF. In Figure 7, the first row shows the input images, the second row shows the ground truth labels for the images, the third row shows the segmented output of the encoder-decoder network without CRF and the last row shows the segmented output of the encoder-decoder network when combined with the CRF module. The output result from Figure 7 shows that; the CRF-based encoder-decoder network approach gives better detection and segmentation performance results than only the encoder-decoder approach on all groups of skin lesions. Both perform better than the traditional FCN method and other existing methods. The reason for the better performance is that the CRF method facilitates feature learning of fine-grained lesions which gives a well-defined lesion boundary as seen in Figure 7 with some improvements in accuracy and dice-coefficient score as shown in the chart in Figure 6. The CRF-based approach outperforms the encoder-decoder network only, which validates the effectiveness of the localization. This also highlights that the combination with the probabilistic graphical CRF model produces segmentation output with more precise borders as shown in the fourth row.

## 2) CLASSIFICATION RESULTS

The classification model was trained and validated on the HAM10000 which is composed of 10030 skin images with corresponding class labels. The dataset is composed of 7 important diagnostic categories of skin lesions which are represented by AKIEC, BCC, BKL, DF, MEL, NV and VASC. Sample skin lesion images from PH2 and ISBI 2017 were also used to test the classification model. For a general evaluation of the classification system, we first evaluated the performance using metrics such as Accuracy, loss, and dice-coefficient. Figure 8 and Figure 9 show the accuracy-loss curve and the dice-coefficient curve of the classification system. From Figure 8, we got the overall accuracy of 98.3% and training loss of 0.6%, and from Figure 9 we got an overall dice-coefficient of 92%. The classification system was evaluated using performance metrics such as accuracy, precision, recall, and F1-score and the results were compared against existing methods such as Deep Convolutional network with transfer learning, Multi-level Densenet, Dilated VGG16 and Dilated InceptionV3 as summarised in Table 3. The classification system outperforms the existing methods with the overall accuracy, precision, recall, and F1-Score

of 98.3%, 98%, 98.5%, and 98.0% respectively when evaluated on HAM10000 as shown in Table 3. The classification results from Table 3 can be analyzed as follows: First, our FCN-based DenseNet classification system obtained the highest overall accuracy of 98.3% when compared with recent deep learning methods (i.e., Multi-level Densenet, Deep Convolution Network with Transfer Learning, Dilated VGG16 and Dilated InceptionV3), indicating better learning ability which is beneficial for skin lesion classification. Second, the FCN-based DenseNet network consistently outperformed the other six deep learning methods in F1-Score, which implies its ability for effective analysis of discriminative features for automatic skin lesion classification. Third, the FCN-based DenseNet network yielded the best performance in all other metrics such as Precision and Recall showing its ability in effective identification of relevant instances and higher measure of completeness and exactness.

The detailed results and experiments focus on comparison of the performance of the classification model on the 7-class (AKIEC, BCC, BKL, DF, MEL, NV, VASC). In order to achieve this, we evaluated the performance using confusion matrix, ROC curve, image classification output, and classification reports in Figure 10, Figure 11, Figure 12, Table 4, Table 5 and Figure 13 respectively. The confusion matrix was reported across all classes for better evaluation of the performance per class and following this, the results of our 7-class predictions were also reported using the ROC curve. The results of the performance analyses were presented through the confusion matrix to get explanatory insights into the results as shown in Figure 10 and Table 5. The following analyses were carried out from the confusion matrix table and reported: 330 AKIEC images were utilized for the experiment; the prediction of 319 images of AKIEC were correctly classified as AKIEC with a classification accuracy of 96.66%. Also, there was a prediction of 2 images of AKIEC FIGURE 8. Accuracy and loss performance curves of the proposed classification model for both validation and training on HAM10000 classified incorrectly as BCC, prediction of 3 images of AKIEC were incorrectly classified as BKL and prediction of 3 images of AKIEC were incorrectly classified as melanoma. 514 BCC images were also used for the experiment; the prediction of 495 images of BCC were correctly classified as BCC with a classification accuracy of 96.30%. Also, there was prediction of 12 images of BCC classified incorrectly as AKIEC, and prediction of 7 images of BCC were incorrectly classified as BKL. 1099 BKL images were also utilized; the prediction of 1095 images of BKL was correctly classified as BKL with a classification accuracy of 99.63%. Also, there was a prediction of 1 image of a BKL classified incorrectly as a Mel, and prediction of 3 images of BKL incorrectly classified as NV. 115 DF images were utilized; the prediction of 111 images of a DF were correctly classified as DF with classification accuracy of 96.52%. Also, there was a prediction of 2 images of a DF classified incorrectly as AKIEC, and prediction of 2 images of a DF were incorrectly classified as BKL. 1112 MEL

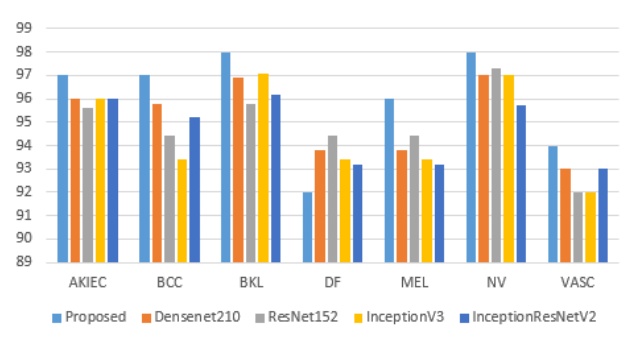
images were utilized for the experiment; the prediction of 1048 images of MEL was correctly classified as MEL with a classification accuracy of 94.24%. Also, there was a prediction of 1 image of an AKIEC classified incorrectly as a MEL, prediction of 18 images of a MEL were incorrectly classified as a BKL and prediction of 45 images of a MEL were incorrectly classified as NV. The total number of 6722 NV images were utilized; the prediction of 6654 images of NV were correctly classified as NV with classification accuracy of 98.98%. Also, there was a prediction of 2 images of an NV classified incorrectly as an AKIEC, prediction of 1 image of an NV incorrectly classified as a BCC, prediction of 53 images of NV were incorrectly classified as BKL, and prediction of 12 images of NV were incorrectly classified as MEL. The total number of 142 VASC images were used; the prediction of 141 images of a VASC were correctly classified as VASC with a classification accuracy of 99.29%. Also, there was a prediction of 1 image of VASC classified incorrectly as an NV.

The results of ROC analysis are illustrated for the 7 classes in Figure 11. Here, we represented AKIEC, BCC, BKL, DF, MEL, NV, VASC by class 0, class 1, class 2, class 3, class 4, class 5, and class 6 respectively. Looking at the ROC curves from Figure 11, the classification system achieved the highest classification performance on the identification of class 2 and class 5 which are BCC and NV with AUC 98%. However, the minimum AUC achieved was 92% for class 3. The overall AUC of this model was 98% with the macro average AUC of 98% and micro-average AUC of 96%. This indicates better detection of BCC and NV classes with the poorest detection performance of BKL which was as a result of the size of their training dataset.

For more insight into the classification performance, Tables 4–5 shows the classification report of the classification results by the 7 classes, respectively.

For more insight into the classification performance, Tables 4–5 shows the classification report of the classification results by the 7 classes, respectively. The classification result and output of the sample images used for testing the proposed classification system are shown in Figure 12. The system was able to identify most images accurately as illustrated in Figure 12 except in only two cases highlighted with the red line: the case of a BCC image classified incorrectly as BKL in the first row and another case of a MEL image classified incorrectly as BKL in the second row. This is further collaborated by the classification performance reports in Table 4 showing the precision, recall, and F1-Score for the 7 diagnostic categories of pigmented skin lesions with MEL and BCC having the lowest recall score of 94% and 96% respectively and also BKL having the lowest precision of 93%.

For the purpose of this work, over the entire 7-class dataset was benchmarked to compare the performance of the state-of-the-arts with the proposed classification method as shown in Figure 13. The proposed method achieves the

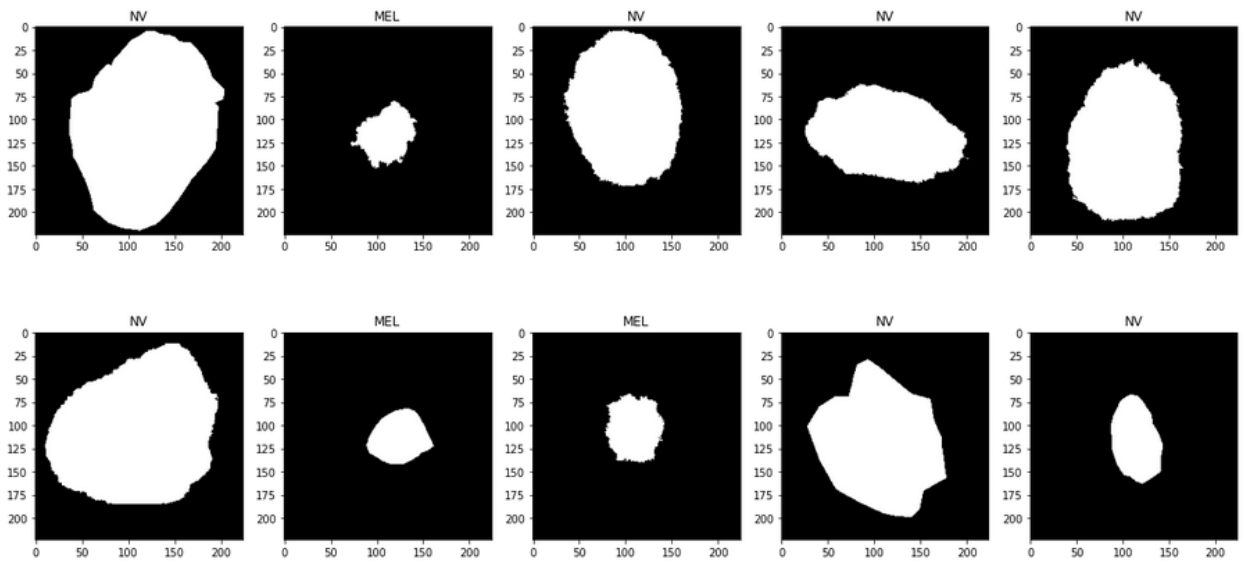


**FIGURE 13.** The figure shows the comparison of the performance of the proposed classification method and other methods on classification of the seven categories of skin lesion.

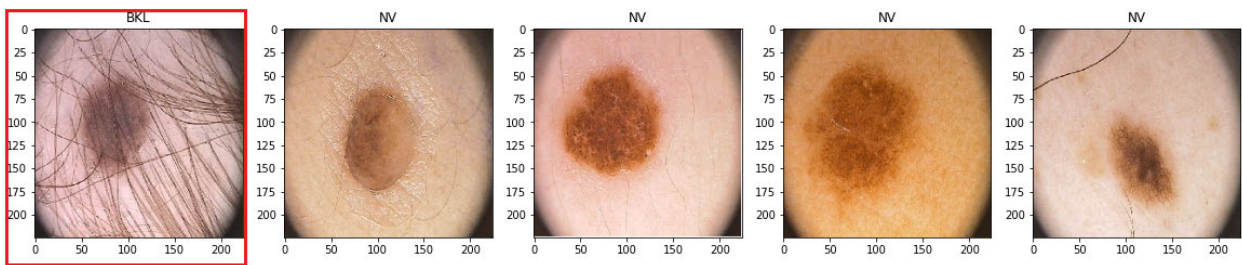
best performance in all the classes when compared with the state-of-the-arts.

**Classification of the segmented images:** The effect of the segmentation process was examined by applying the classification networks to classify samples of well-segmented images from the ISBI 2017 Dataset. The classification result of segmented images in Figure 14 shows that the skin lesion images were all correctly classified as either NV or MEL. This shows the effect of segmentation on the performance of the proposed classification model. There was improvement suggesting that, better performance of the classifier on the segmented images than its performance on the unsegmented images. This leverages on the effect of CRF-based multi-scale encoder-decoder network to effectively pre-process the images and improve the detection accuracy, especially for the skin lesions with complex features. The FCN-based Densenet network optimized with CRF-based multi-scale encoder-decoder network yielded better performance compared with only the FCN-based Densenet network and the existing state-of-the-art. For example, the overall accuracy was improved by 0.6 (98.90 vs. 98.30), precision improved by 0.5 (98.5 vs. 98.0), recall improved by 0.5 (99 vs. 98.5) and F1-score improved by 0.5 (98.5 vs. 98.0) respectively as shown in Figure 16. It indicates that incorporating the CRF-based multi-scale encoder-decoder network could effectively refine the feature learning to improve the lesion detection accuracy. This is also discovered in Figure 14 where all the 10 images were detected correctly with 98.9% accuracy.

**Generalization Effects:** We also tested the generalization performance of the FCN-based Densenet network on the classification of sample skin lesion images from another skin lesion dataset, PH2, as shown in Figure 15. All the images were correctly classified except the first image that was wrongly classified as BKL whereas the image is MEL. This also shows the effect of segmentation because the sample images used were not preprocessed and segmented. Finally, the effect of segmentation also reflects on the dice-coefficient curve in Figure 9 where the images used were un-segmented.



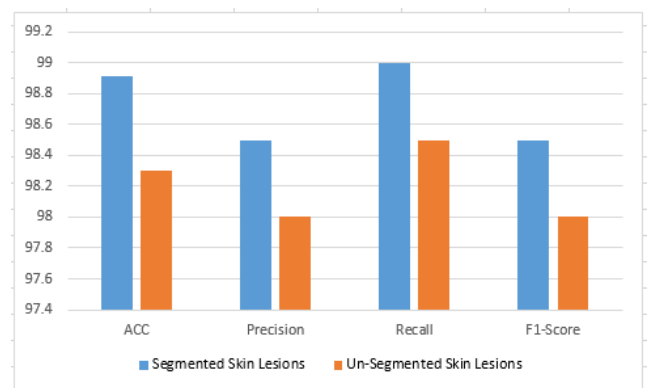
**FIGURE 14.** The figure shows the classification results of the proposed classification method on some sample segmented skin lesion images.



**FIGURE 15.** The figure shows the classification result output of the proposed classification method on some sample PH2 skin lesion images.

### 3) EFFECTS OF HYPER-PARAMETERS TUNING

We performed extensive experiments for hyper-parameters tuning to achieve optimal performance for the proposed system. Some sets of experiments were conducted to show the effects of tuning of these hyper-parameters as shown in Table 6. Major hyper-parameters in the network such as learning rate, optimizer, decay constant, and the number of dense layers as were varied and tuned. Three major optimization algorithms were explored and tuned which include Adaptive Moment Optimization (Adam), stochastic gradient descent algorithm (SGD) and Root Mean Square Propagation (RMSprop) optimizers. The aim is to reduce overfitting and make better predictions with the model. Experiment results are presented in Table 6 which shows the impact of varying these hyper-parameters on the system performance. From the results, the first row presents the best result in terms of accuracy. The experiment results show that the accuracy of the model performance has a significant improvement by using hyper-parameter optimization algorithms. In order to achieve optimal performance, the major hyper-parameters were explored. These parameters were varied as follows:



**FIGURE 16.** The figure compares the general performance (AUC) of the proposed classification method on segmented and un-segmented lesion images.

- 1) **Optimizers:** We considered and varied three most general optimization algorithms: ADAM, SGD and RMSPROP. The optimization algorithms generally affect the training speed and the final predictive performance of deep learning models. We performed an experiment in which all the optimization



**TABLE 6.** Table showing the impact of varying the Hyper-Parameters on the Performance of the Proposed System.

Optimizers	Learning Rate	Weight Decay Values	Dense Layers Level	Time/epoch(s)	Accuracy	F1-Score	Recall
<b>Adam</b>	<b>0.0001</b>	<b>0.001</b>	<b>6</b>	<b>802</b>	<b>98.0</b>	<b>98.0</b>	<b>98.5</b>
Adam	0.0001	0.0001	6	809	97.67	95.42	94.30
Adam	0.0001	0.01	6	805	95.06	92.03	90.12
Adam	0.001	0.0001	6	801	96.50	94.45	93.69
Adam	0.001	0.001	6	802	95.50	93.05	92.15
Adam	0.001	0.01	6	801	94.70	92.03	90.69
Adam	0.01	0.0001	6	801	95.50	93.45	93.69
Adam	0.01	0.001	6	802	93.50	92.05	91.15
Adam	0.01	0.01	6	801	92.70	91.03	90.69
Adam	0.0001	0.0001	4	802	96.86	94.03	90.12
Adam	0.0001	0.001	4	809	95.27	94.42	87.30
Adam	0.0001	0.01	4	802	94.06	91.03	86.12
Adam	0.001	0.0001	4	801	95.50	93.45	90.69
Adam	0.001	0.001	4	802	94.50	92.05	86.15
Adam	0.001	0.01	4	801	93.70	91.03	85.69
Adam	0.01	0.0001	4	801	94.50	92.45	90.69
Adam	0.01	0.001	4	802	91.50	91.05	84.15
Adam	0.01	0.01	4	801	90.70	90.03	83.69
SGD	0.0001	0.001	6	809	97.67	95.42	90.30
SGD	0.0001	0.0001	6	809	96.67	93.42	89.30
SGD	0.0001	0.01	6	802	95.06	92.03	86.12
SGD	0.001	0.0001	6	801	96.50	94.45	87.69
SGD	0.001	0.001	6	802	95.50	93.05	86.15
SGD	0.001	0.01	6	801	94.70	92.03	85.69
SGD	0.01	0.0001	6	801	95.50	93.45	88.69
SGD	0.01	0.001	6	802	93.50	92.05	84.15
SGD	0.01	0.01	6	801	92.70	91.03	83.69
SGD	0.0001	0.0001	4	802	96.86	94.73	89.10
SGD	0.0001	0.001	4	809	95.27	92.22	86.10
SGD	0.0001	0.01	4	802	94.16	91.93	85.2
SGD	0.001	0.0001	4	801	95.80	93.5	86.29
SGD	0.001	0.001	4	802	94.10	92.0	85.10
SGD	0.001	0.01	4	801	93.20	91.83	84.19
SGD	0.01	0.0001	4	801	94.10	92.65	84.29
SGD	0.01	0.001	4	802	92.20	91.45	83.55
SGD	0.01	0.01	4	801	91.20	90.03	81.29
RMSprop	0.0001	0.0001	6	809	98.67	93.42	92.30
RMSprop	0.0001	0.001	6	809	97.08	95.42	88.30
RMSprop	0.0001	0.01	6	802	94.06	92.03	86.12
RMSprop	0.001	0.0001	6	801	97.50	94.45	91.69
RMSprop	0.001	0.001	6	802	96.50	93.75	90.15
RMSprop	0.001	0.01	6	801	95.70	92.90	88.69
RMSprop	0.01	0.0001	6	801	96.50	93.45	90.69
RMSprop	0.01	0.001	6	802	94.50	92.05	88.15
RMSprop	0.01	0.01	6	801	93.70	91.03	87.69
RMSprop	0.0001	0.0001	4	802	97.06	95.03	90.12
RMSprop	0.0001	0.001	4	809	96.27	95.42	87.30
RMSprop	0.0001	0.01	4	802	93.06	92.03	86.12
RMSprop	0.001	0.0001	4	801	96.80	94.05	90.90
RMSprop	0.001	0.001	4	802	94.80	93.05	88.15
RMSprop	0.001	0.01	4	801	93.90	92.03	85.69
RMSprop	0.01	0.0001	4	801	95.50	93.15	90.0
RMSprop	0.01	0.001	4	802	93.50	92.05	87.15
RMSprop	0.01	0.01	4	801	92.70	90.03	86.69

hyper-parameters were varied for each optimizer. The performance of these optimization algorithms were compared as shown in Table 6 after tuning their respective hyper-parameters with the Adam optimizer produced the best performance followed by SGD and then RMSprop.

2) **Learning Rate:** We experimented between three values of learning rates: 0.01, 0.001, and 0.0001.

We achieved the best system performance with a small learning rate value of 0.0001. Model training tends to diverge when the learning rates become too large. A decreased learning rate yielded an improved generalization accuracy for the proposed model.

3) **Weight Decay Constant:** We experimented between three values of weight decay constants: 0.01, 0.001 and 0.0001. The weight decay value of 0.01 eventually

produced the best system performance with Adam optimizer. This can be due to the nature of our dataset and the architecture which are not too complex.

- 4) **Dropout rate:** We also utilized dropout as a regularization technique to avoid over-fitting and increase the validation accuracy and thus increasing the generalizing power. We experimented between 0.5 and 0.25 values. The value 1.0 means no dropout, and 0.0 means no outputs from the layer.
- 5) **Number of Dense Layers:** In order to reduce the complexity of the architecture, we experimented between dense layers number 4 and 6 number of dense layers for the system architecture. We achieved optimal performance at level 6.
- 6) **Batch size:** Recent empirical research [81] has shown that increasing the batch size also affects the training speed. In this research, the batch size of 128 produced the best results.

**Finalizing Hyper-parameters Setting:** The system achieved the best performance by setting the hyper-parameters as stated below and in Table 6:

- 1) Optimizers: Adam
- 2) Learning Rate: 0.0001
- 3) Weight Decay Values: 0.001
- 4) Dense Layers Level: 6

With these settings, we were able to achieve the optimal performance for the proposed system.

Lastly, it can be stated that the proposed light-weight classification system achieved better performance with reduced computing resources that can meet up with the requirement in the real-time clinical practice. The system performed better than most existing methods and can meet up with real time medical diagnosis task in diagnosing skin cancer with the processing time for each dermoscopy image at averagely 8s. The performance evaluation of our model was done under the same hardware conditions and the same dataset with some state-of-the-arts as shown in Table 7 and the comparison result shows that the proposed system outperforms these existing techniques in the computational speed during both the training and testing phases.

**TABLE 7. Measurements in seconds of the training time per epoch and test time per skin lesion image.**

Techniques	Average Training Time/ epoch	Test Time/Image
Proposed model	800	8
ResNet152	950	12
InceptionV3	1050	15
VGG19	1600	23

## V. CONCLUSION

This work provides some novel approaches using deep learning techniques in the segmentation and classification methodologies of skin lesion images towards detection and diagnosis of skin cancer. A deep learning-based CAD framework that is composed of a multi-scale encoder-decoder segmentation network and an FCN-based DenseNet classification network,

has been proposed for the detection and classification of skin lesion images to diagnose skin cancer diseases. The proposed method was evaluated on publicly available database of HAM 10000 that is made up of 7 important diagnostic categories of skin lesion and it has shown superior performance than the existing state-of-the-art methods in most of the classification performance evaluation; and most especially in both the segmentation and classification accuracies. The system includes a segmentation stage which employs a novel encoder-decoder network that is integrated into CRF module for accurate and refined lesion border detection. It also includes an FCN-based DenseNet framework for an efficient classification of skin lesions. This was shown to outperform existing state-of-the-arts classification techniques. It was established that introducing the multi-scale encoder-decoder segmentation network into the classification system will improve the classification accuracy of the entire system. The classification system was evaluated separately on unsegmented images to show the effect of the segmentation network. It can be concluded from our results that application of efficient pre-processing and segmentation techniques on skin lesion images before classification can lead to better detection performance of deep learning-based classification system. The proposed system has been able to overcome the challenges of dealing with the complex features of skin lesion images and heavy parameter tuning of the traditional CNN.

## REFERENCES

- [1] A. R. D. Delbridge, L. J. Valente, and A. Strasser, "The role of the apoptotic machinery in tumor suppression," *Cold Spring Harbor Perspect. Biol.*, vol. 4, no. 11, Nov. 2012, Art. no. a008789
- [2] M. A. Albahar, "Skin lesion classification using convolutional neural network with novel regularizer," *IEEE Access*, vol. 7, pp. 38306–38313, 2019.
- [3] R. Marks, "An overview of skin cancers," *Cancer*, vol. 75, no. S2, pp. 607–612, Jan. 1995.
- [4] H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer (Keratinocyte Carcinomas) in the US population, 2012," *JAMA Dermatology*, vol. 151, no. 10, p. 1081, Oct. 2015.
- [5] *Cancer Facts and Figures 2017*. Accessed: Jan. 20, 2020. [Online]. Available: <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures.html>
- [6] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics," *CA, A cancer J. Clinicians*, vol. 65, no. 1, pp. 5–29, 2015.
- [7] S. A. Gandhi and J. Kampp, "Skin cancer epidemiology, detection, and management," *Med. Clinics North Amer.*, vol. 99, no. 6, pp. 1323–1335, Nov. 2015.
- [8] W. E. Damsky and M. Bosenberg, "Melanocytic nevi and melanoma: Unraveling a complex relationship," *Oncogene*, vol. 36, no. 42, pp. 5771–5792, Oct. 2017.
- [9] N. Eisemann, A. Waldmann, A. C. Geller, M. A. Weinstock, B. Volkmer, R. Greinert, E. W. Breitbart, and A. Katalinic, "Non-melanoma skin cancer incidence and impact of skin cancer screening on incidence," *J. Investigative Dermatology*, vol. 134, no. 1, pp. 43–50, Jan. 2014.
- [10] M. Thorn, F. Ponte, R. Bergstrom, P. Sparen, and H.-O. Adami, "Clinical and histopathologic predictors of survival in patients with malignant melanoma: A population-based study in sweden," *JNCI J. Nat. Cancer Inst.*, vol. 86, no. 10, pp. 761–769, May 1994.
- [11] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
- [12] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *Lancet Oncol.*, vol. 3, no. 3, pp. 159–165, Mar. 2002.

- [13] G. Argenziano, H. P. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, M. Binder, "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet," *J. Amer. Acad. Dermatology*, vol. 5, pp. 679–693, May 2003.
- [14] M. Binder, M. Schwarz, A. Winkler, A. Steiner, A. Kaider, K. Wolff, and H. Pehamberger, "Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists," *Arch. Dermatology*, vol. 131, no. 3, pp. 286–291, Mar. 1995.
- [15] J. Gachon, P. Beaulieu, J. F. Sei, J. Gouvernet, J. P. Claudel, M. Lemaitre, M. A. Richard, and J. J. Grob, "First prospective study of the recognition process of melanoma in dermatological practice," *Arch. Dermatology*, vol. 141, no. 4, pp. 434–438, Apr. 2005.
- [16] A. R. Ratul, M. H. Mozaffari, W.-S. Lee, and E. Parimbelli, "Skin lesions classification using deep learning based on dilated convolution," *BioRxiv*, Jan. 2019, Art. no. 860700.
- [17] V. Revathi and A. Chithra, "A review on segmentation techniques in skin lesion images," *Intl Res. J. Eng. Tech. (IRJET)*, vol. 2, no. 9, pp. 2598–2603, 2015.
- [18] Q. Abbas, I. F. Garcia, M. Emre Celebi, W. Ahmad, and Q. Mushtaq, "A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images," *Skin Res. Technol.*, vol. 19, no. 1, pp. e490–e497, Feb. 2013.
- [19] A. A. Adegun and S. Viriri, "Deep learning-based system for automatic melanoma detection," *IEEE Access*, vol. 8, pp. 7160–7172, 2020.
- [20] H. Alquran, I. A. Qasmieh, A. M. Alqudah, S. Alhammouri, E. Alawneh, A. Abughazaleh, and F. Hasayen, "The melanoma skin cancer detection and classification using support vector machine," in *Proc. IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol. (AEECT)*, Oct. 2017, pp. 1–5.
- [21] K. Hoffmann *et al.*, "Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy," *Brit. J. Dermatology*, vol. 149, no. 4, pp. 801–809, Oct. 2003.
- [22] N. Hameed, F. Hameed, A. Shabut, S. Khan, S. Cirstea, and A. Hossain, "An intelligent computer-aided scheme for classifying multiple skin lesions," *Computers*, vol. 8, no. 3, p. 62, Aug. 2019.
- [23] A. Murugan, S. A. H. Nair, and K. P. S. Kumar, "Detection of skin cancer using SVM, random forest and kNN classifiers," *J. Med. Syst.*, vol. 43, no. 8, p. 269, Aug. 2019.
- [24] R. D. Seeja and A. Suresh, "Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM)," *Asian Pacific J. Cancer Prevention*, vol. 20, no. 5, pp. 1555–1561, May 2019.
- [25] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *Sensors*, vol. 18, no. 2, p. 556, Feb. 2018.
- [26] S. M. Rajpara, A. P. Botello, J. Townend, and A. D. Ormerod, "Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma," *Brit. J. Dermatology*, vol. 161, no. 3, pp. 591–604, Sep. 2009.
- [27] A. Hekler, J. S. Utikal, A. H. Enk, A. Hauschild, M. Weichenthal, R. C. Maron, C. Berking, S. Haferkamp, J. Klode, D. Schadendorf, B. Schilling, T. Holland-Letz, B. Izar, C. Von Kalle, S. Fröhling, and T. J. Brinker, "Superior skin cancer classification by the combination of human and artificial intelligence," *Eur. J. Cancer*, vol. 120, pp. 114–121, Oct. 2019.
- [28] T. J. Brinker, A. Hekler, J. S. Utikal, N. Grabe, D. Schadendorf, J. Klode, C. Berking, T. Steeb, A. H. Enk, and C. von Kalle, "Skin cancer classification using convolutional neural networks: Systematic review," *J. Med. Internet Res.*, vol. 20, no. 10, Oct. 2018, Art. no. e11936.
- [29] S. R. Guha and S. R. Haque, "Performance comparison of machine learning-based classification of skin diseases from skin lesion images," in *Proc. Int. Conf. Commun., Comput. Electron. Syst.*, Singapore: Springer, 2020, pp. 15–25.
- [30] L. Bi, D. Feng, and J. Kim, "Dual-path adversarial learning for fully convolutional network (FCN)-based medical image segmentation," *Vis. Comput.*, vol. 34, nos. 6–8, pp. 1043–1052, Jun. 2018.
- [31] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, "Dermoscopic image segmentation via multistage fully convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2065–2074, Sep. 2017.
- [32] B. Abdollahi, N. Tomita, and S. Hassanpour, "Data Augmentation in Training Deep Learning Models for Medical Image Analysis," in *Deep Learners and Deep Learner Descriptors for Medical Applications*. Cham, Switzerland: Springer, 2020, pp. 167–180.
- [33] K. Raza and N. Kumar Singh, "A tour of unsupervised deep learning for medical image analysis," 2018, *arXiv:1812.07715*. [Online]. Available: <http://arxiv.org/abs/1812.07715>
- [34] A. Vouliodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, Feb. 2018.
- [35] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, "Step-wise integration of deep class-specific learning for dermoscopic image segmentation," *Pattern Recognit.*, vol. 85, pp. 78–89, Jan. 2019.
- [36] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Commun. ACM*, vol. 54, no. 10, pp. 95–103, Oct. 2011.
- [37] S. Pereira, R. Meier, R. McKinley, R. Wiest, V. Alves, C. A. Silva, and M. Reyes, "Enhancing interpretability of automatically extracted machine learning features: Application to a RBM-random forest system on brain lesion segmentation," *Med. Image Anal.*, vol. 44, pp. 228–244, Feb. 2018.
- [38] M. Akhavan Aghdam, A. Sharifi, and M. M. Pedram, "Combination of rs-fMRI and sMRI data to discriminate autism spectrum disorders in young children using deep belief network," *J. Digit. Imag.*, vol. 31, no. 6, pp. 895–903, Dec. 2018.
- [39] A. Al Nahid, A. Mikaelian, and Y. Kong, "Histopathological breast-image classification with restricted Boltzmann machine along with backpropagation," *Biomed. Res.*, vol. 29, no. 10, pp. 2068–2077, 2018.
- [40] Y. Zhu, L. Wang, M. Liu, C. Qian, A. Yousuf, A. Oto, and D. Shen, "MRI-based prostate cancer detection with high-level representation and hierarchical classification," *Med. Phys.*, vol. 44, no. 3, pp. 1028–1039, Mar. 2017.
- [41] S. Minaee, Y. Wang, A. Aygar, S. Chung, X. Wang, Y. W. Lui, E. Fieremans, S. Flanagan, and J. Rath, "MTBI identification from diffusion MR images using bag of adversarial visual features," *IEEE Trans. Med. Imag.*, vol. 38, no. 11, pp. 2545–2555, Nov. 2019.
- [42] V. M. Vergara, A. R. Mayer, E. Damaraju, K. A. Kiehl, and V. Calhoun, "Detection of mild traumatic brain injury by machine learning classification using resting state functional network connectivity and fractional anisotropy," *J. Neurotrauma*, vol. 34, no. 5, pp. 1045–1053, Mar. 2017.
- [43] A.-R. Ali, J. Li, and T. Trappenberg, "Supervised versus unsupervised deep learning based methods for skin lesion segmentation in dermoscopy images," in *Proc. Can. Conf. Artif. Intell.*, Cham, Switzerland: Springer, 2019, pp. 373–379.
- [44] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," 2020, *arXiv:2001.05566*. [Online]. Available: <http://arxiv.org/abs/2001.05566>
- [45] S. Feng, Z. Zhuo, D. Pan, and Q. Tian, "CcNet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features," *Neurocomputing*, vol. 392, pp. 268–276, Jun. 2020.
- [46] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Apr. 1, 2020, doi: [10.1109/TPAMI.2020.2983686](https://doi.org/10.1109/TPAMI.2020.2983686).
- [47] S. Vesal, S. M. Patil, N. Ravikumar, and A. K. Maier, "A multi-task framework for skin lesion detection and segmentation," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Cham, Switzerland: Springer, 2018, pp. 285–293.
- [48] N. Hameed, A. M. Shabut, M. K. Ghosh, and M. A. Hossain, "Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques," *Expert Syst. Appl.*, vol. 141, Mar. 2020, Art. no. 112961.
- [49] A.-R. Ali, J. Li, G. Yang, and S. J. O'Shea, "A machine learning approach to automatic detection of irregularity in skin lesion border using dermoscopic images," *PeerJ Comput. Sci.*, vol. 6, Jun. 2020, Art. no. e268.
- [50] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie, "Sample-efficient deep learning for COVID-19 diagnosis based on CT scans," *MedRxiv*, vol. 7, Jan. 2020.
- [51] H. El-Khatib, D. Popescu, and L. Ichim, "Deep learning-based methods for automatic diagnosis of skin lesions," *Sensors*, vol. 20, no. 6, p. 1753, Mar. 2020.
- [52] N. Hameed, A. M. Shabut, and M. A. Hossain, "Multi-class skin diseases classification using deep convolutional neural network and support vector machine," in *Proc. 12th Int. Conf. Softw., Knowl., Inf. Manage. Appl. (SKIMA)*, Dec. 2018, pp. 1–7.

- [53] J.-A. Almaraz-Damian, V. Ponomaryov, S. Sadovnychiy, and H. Castillejos-Fernandez, "Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures," *Entropy*, vol. 22, no. 4, p. 484, Apr. 2020.
- [54] S. Kalouche, A. Ng, and J. Duch, "Vision-based classification of skin cancer using deep learning," in *Conducted on Stanfords Machine Learning Course (CS)*. Stanford, CA, USA: Stanford Univ., 2015.
- [55] A. Soudani and W. Barhoumi, "An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction," *Expert Syst. Appl.*, vol. 118, pp. 400–410, Mar. 2019.
- [56] K. M. Hosny, M. A. Kassem, and M. M. Foad, "Classification of skin lesions using transfer learning and augmentation with alex-net," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0217293.
- [57] T. Akram, H. M. J. Lodhi, S. R. Naqvi, S. Naem, M. Alhaisoni, M. Ali, S. A. Haider, and N. N. Qadri, "A multilevel features selection framework for skin lesion classification," *Human-centric Comput. Inf. Sci.*, vol. 10, no. 1, pp. 1–26, Dec. 2020.
- [58] B. Ahmad, M. Usama, C.-M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, "Discriminative feature learning for skin disease classification using deep convolutional neural network," *IEEE Access*, vol. 8, pp. 39025–39033, 2020.
- [59] M. A. Al-masni, D.-H. Kim, and T.-S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Comput. Methods Programs Biomed.*, vol. 190, Jul. 2020, Art. no. 105351.
- [60] A. Naem, M. S. Farooq, A. Khelifi, and A. Abid, "Malignant melanoma classification using deep learning: Datasets, performance measurements, challenges and opportunities," *IEEE Access*, vol. 8, pp. 110575–110597, 2020.
- [61] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefler, "Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data," *MethodsX*, vol. 7, Mar. 2020, Art. no. 100864.
- [62] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Comput. Methods Programs Biomed.*, vol. 162, pp. 221–231, Aug. 2018.
- [63] K. Shimizu, H. Iyatomi, M. E. Celebi, K.-A. Norton, and M. Tanaka, "Four-class classification of skin lesions with task decomposition strategy," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 274–283, Jan. 2015.
- [64] A. M. Alqudah, H. Alquraan, and 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. Biomimetics, Biomater. Biomed. Eng.*, vol. 42, pp. 67–78, Jul. 2019.
- [65] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1397–1400.
- [66] G. Zhang, X. Shen, S. Chen, L. Liang, Y. Luo, J. Yu, and J. Lu, "DSM: A deep supervised multi-scale network learning for skin cancer segmentation," *IEEE Access*, vol. 7, pp. 140936–140945, 2019.
- [67] A. Adegun and S. Viriri, "An enhanced deep learning framework for skin lesions segmentation," in *Proc. Int. Conf. Comput. Collective Intell.*, Cham, Switzerland: Springer, 2019, pp. 414–425.
- [68] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Mach. Learn. (ICML)*. Philadelphia, PA, USA: Univ. of Pennsylvania, 2001, pp. 282–289. [Online]. Available: <http://portal.acm.org/citation.cfm?id=655813>
- [69] K. Philipp and V. Koltun, "Efficient inference in fully connected CEFs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 109–117.
- [70] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, Dec. 2018, Art. no. 180161.
- [71] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kallou, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," 2019, *arXiv:1902.03368*. [Online]. Available: <http://arxiv.org/abs/1902.03368>
- [72] *PH2 Database*. Accessed: Jan. 20, 2020. [Online]. Available: <https://www.fc.up.pt/addi/ph2database.html>
- [73] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kallou, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," 2017, *arXiv:1710.05006*. [Online]. Available: <http://arxiv.org/abs/1710.05006>
- [74] F. Xie, J. Yang, J. Liu, Z. Jiang, Y. Zheng, and Y. Wang, "Skin lesion segmentation using high-resolution convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 186, Apr. 2020, Art. no. 105241.
- [75] Ş. Öztürk and U. Özkaya, "Skin lesion segmentation with improved convolutional neural network," *J. Digit. Imag.*, May 2020, doi: [10.1007/s10278-020-00343-z](https://doi.org/10.1007/s10278-020-00343-z).
- [76] S. Mishra, P. Liang, A. Czajka, D. Z. Chen, and X. S. Hu, "CC-NET: Image complexity guided network compression for biomedical image segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 57–60.
- [77] Zhang, Zhenli, Xiangyu Zhang, Chao Peng, Xiangyang Xue, and Jian Sun, "Exfuse: Enhancing feature fusion for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 269–284, 2018.
- [78] A. Rezvantalab, H. Safigholi, and S. Karimijeshni, "Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms," 2018, *arXiv:1810.10348*. [Online]. Available: <http://arxiv.org/abs/1810.10348>
- [79] M. A. Kassem, K. M. Hosny, and M. M. Foad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020.
- [80] P. Carcagni, M. Leo, A. Cuna, P. L. Mazzeo, P. Spagnolo, G. Celeste, and C. Distanto, "Classification of skin lesions by combining multilevel learnings in a DenseNet architecture," in *Proc. Int. Conf. Image Anal. Process.*, Cham, Switzerland: Springer, 2019, pp. 335–344.
- [81] G. Zhang, L. Li, Z. Nado, J. Martens, S. Sachdeva, G. Dahl, C. Shallue, and R. B. Grosse, "Which algorithmic choices matter at which batch sizes? insights from a noisy quadratic model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8196–8207.



**ADEKANMI A. ADEGUN** received the B.Tech. degree in computer science from the Ladoko Akintola University of Technology (LAUTECH), Ogbomosho, and the M.Sc. degree in computer science from the University of Ibadan, Nigeria. He is currently pursuing the Ph.D. degree in computer science with the University of KwaZulu-Natal, South Africa. He is involved in developing deep learning models for medical image analysis at the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal. He is also a Lecturer at the Department of Computer Science, Landmark University, Nigeria. He has published some of his research works in some accredited computer vision and image processing journals and international and national conference proceedings. His research interests include machine learning, computer vision, image processing, pattern recognition, and medical image analysis.



**SERESTINA VIRIRI** (Member, IEEE) is currently a Professor of computer sciences. He is also a Professor with the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, South Africa. He is an NRF-Rated Researcher. He has published extensively in several accredited computer vision and image processing journals and international and national conference proceedings. His research interests include computer vision, image processing, pattern recognition, and other image processing related areas, such as bio-metrics, medical imaging, and nuclear medicine. He serves as a reviewer for several accredited journals. He has also served on the program committees for numerous international and national conferences. He has graduated several M.Sc. and Ph.D. students.

...