Stacked Cross Validation with Deep Features: A Hybrid Method for Skin Cancer Detection

Ahmed Al-Karawi, Ercan Avşar*

Abstract: Detection of malignant skin lesions is important for early and accurate diagnosis of skin cancer. In this work, a hybrid method for malignant lesion detection from dermoscopy images is proposed. The method combines the feature extraction process of convolutional neural networks (CNN) with an ensemble learner called stacked cross-validation (CV). The features extracted by three different CNN architectures, namely, ResNet50, Xception, and VGG16 are used for training of four different baseline classifiers, which are support vector machines, k-nearest neighbors, artificial neural networks, and random forests. The stacked outputs of these classifiers are used to train a logistic regression model as a meta-classifier. The performance of the proposed method is compared with the baseline classifiers trained individually as well as AdaBoost classifier, another ensemble learner. Feature extraction with Xception architecture, outperforms all other benchmark models by achieving scores of 0.909, 0.896, 0.886, and 0.917 for accuracy, F1-score, sensitivity, and AUC, respectively.

Keywords: Convolutional Neural Networks; Cross Validation; Deep Learning; Dermoscopy; Skin Cancer; Stacking

1 INTRODUCTION

The medical science is facing the challenging task to detect and cure of cancer in human beings. Skin cancer is the most common cancer type in the United States of America and the most common form of the skin cancer is melanoma [1]. Among different types of skin cancers, malignant melanoma itself causes more than 10,000 deaths annually in the United States [2, 3]. The development of melanoma begins with the production of cysts in the pigment melanin, which is responsible for the color of the skin. It can spread to the lower part of the skin, enter the bloodstream and then spread to other parts of the body. The treatment of this cancer type is a demanding task in its growing stage. Therefore, the early-stage detection of melanoma is important to successfully treat the patient and reduce the rate of mortality.

The early diagnosis of skin cancer is possible through computer-aided devices and tools. The computer-aided diagnostic tools can help the clinicians to improve the clinical diagnosis accuracy for cancer detection. Dermoscopy is the most important non-invasive computer-aided tool for the detection of melanoma as well as the other pigmented skin cancer types [4]. The conventional method for identification of the primary features of melanoma is the eye-based examination of dermoscopy images. These features are the surface structure and the skin color. This examination strategy allows for better differentiation between different cancer types based on their color properties and morphological features [5]. However, visual inspection of dermoscopy images by clinicians relies on expertise and experience. Since human interpretation is a subjective operation, computer-aided intelligent systems are important tools for automatic analysis of the dermoscopy image and reduce the human-related diagnostic errors [6]. For the identification of melanoma, utilization of dermoscopy images together with computer-based tools may improve the diagnostic accuracy of the disease because dermoscopy provides magnified and illuminated images of the skin. As a result, dermoscopy can be considered as a useful tool for computer-based diagnosis systems that implement various methods in image processing, computer vision and machine learning [7].

Nowadays, deep learning has become the most popular and robust technique for various image classification problems. The conventional classification techniques were restricted to the transformation of raw input to formatted features to perform classification task [8]. The deep learning methods allow for automated classification and prediction system because they enable automatic extraction of features from the given images. For performance improvement and high accuracy based results, researchers are turning towards the development of hybrid approaches using deep learning methods [9]. Information regarding the previous studies of skin cancer identification reveals that there is a large scope of studies based on deep neural network approaches. However, there is not sufficient evidence of research on the use of deep learning methods together with stacking algorithms for analyzing dermoscopy images.

In this paper, a hybrid method for classification of dermoscopy images belonging to malignant and benign skin cancer is proposed. The method employs stacked cross-validation (CV) algorithm together with deep neural networks. Convolutional layers of deep neural networks are used for extracting the features from the images, hence they are called as deep features in this work. These features are used to train for different classifiers whose outputs are then stacked to generate a meta-classifier.

The proposed method is named as Stacked Cross-Validation with Deep Features (SCV-DF) and implemented on three levels. At the first level, the deep learning methods are applied to the original dataset to extract features from the images. The outcome of the first level is used as features for the second level, where four different classifiers namely, support vector machines (SVM), k-nearest neighbors (KNN), artificial neural networks (ANN), and random forest (RF) are trained separately. The outcome of the second level is used as features for the third level. The logistic regression method is used in training for third level folds. The output of the level three models is the final prediction and is used as outcome results. The results of the proposed method are compared

with six benchmark models, and it has been shown that the usage of SCV-DF improves the classification performance. Therefore, the main contribution of this work is the proof of concept about the suitability of the deep learning based ensemble models for malignancy detection in skin cancer.

The rest of the paper is structured as follows: Section 2 involves the related work. Details of the experiments as well as the dataset are provided in Section 3. Results and relevant discussion are given Section 4, and Section 5 presents the final conclusions.

2 RELATED WORK

In the past decade, numerous amount of research has been done for the detection and identification of malignant and benign skin cancer. Various methods based on splitting, merging, clustering, and classification was used by the researchers for this task. Each method has some limitations and advantages for the experimental analysis and helps the medical experts in decision making.

The visual properties of the lesions are the most commonly used features for skin cancer identification. These properties are analyzed under three main methods, namely, ABCD method, Seven-point checklist method, and Menzies method. Asymmetry, border, color and diameter are the properties inspected under the ABCD method. In the seven-point checklist method, blue-white veils, atypical pigment and vascular networks, regression structures, and irregularities in globules, blotches and streaks are analyzed. The Menzies method investigates the features of positive and negative lesions by observing the symmetry and color based features. The strategies proposed by these methods are utilized by the researchers to develop computer-based algorithms [10].

One common method for feature extraction of the skin cancer images is the wavelet transform. The texture, border, and geometry features were extracted using the wavelet-decomposition and boundary-series model explained by Rajasekhar et al. [11]. The classification of skin cancer was performed using well-known machine learning algorithms such as SVM, RF, logistic model tree, and hidden Naive Bayes methods. In another method utilizing wavelet transform, feature extraction together with texture analysis were implemented [12]. The extracted features were passed as an input to the stack auto encoders towards classification of malignant and benign skin cancer.

Segmentation methods are also used for feature extraction purposes. A particular segment of the tissue was extracted from melanoma images using watershed segmentation [13]. A measure of asymmetry, border irregularity, color variation, diameter, and texture features were used for the classification of the images. The classification was implemented using KNN, RF and SVM methods. The SVM classifier was found to be robust and dominant when compared to other methods.

There are a number of other works in which traditional machine learning methods like SVM [14, 15], ANN [16, 17] and decision trees [12, 18] are utilized. However, with the advancement of deep learning, various possibilities for skin cancer detection have emerged. One example application may be the segmentation of skin lesions with deep learning methods [19] as well as the classification of the obtained segments [20].

Various convolutional neural network (CNN) architectures are widely used in classification of dermoscopic images. Hekler et al. proposed a method that combines the decisions of humans and ResNet50 models to improve the detection accuracy [21]. A gradient boosting method, XGBoost is used to fuse the decisions and it was shown that this procedure might improve the detection accuracy for some of the classes. On the other hand, S. H. Kassani and P. H. Kassani performed transfer learning on five well known CNN architectures, namely AlexNet, ResNet50, VGG16, VGG19, and Xception [22]. They used a seven-class dataset and reported results with and without data augmentation applied to it. In a study by Codella et al., an ensemble of deep residual network (DRN), CaffeNet and fully convolutional U-Net architectures is proposed for the detection of malignant skin lesions [23]. They used weights for pretrained models of DRN and CaffeNet for feature extraction from the images and showed that generating an ensemble model together with a segmentation step may improve the detection performance. Region-based CNN (RCNN) methods are also utilized for detecting malignant skin lesions. For example, utilization of faster RCNN was proposed by Jianni et al. and it was shown that it outperforms the mean accuracy of the decisions made by dermatologists [24].

As can be seen from the previous works about skin cancer classification, the majority of the researchers either apply an image processing step followed by a machine learning method or feed images of a deep learning model for the training of the model. Therefore, it can be concluded that utilization of ensemble learning techniques together with deep learning models has not been studied in the literature yet. With this study we aim to fill this existing research gap by proposing a novel method for skin cancer detection. In this work, only the convolutional layers of different CNN architectures are used for feature extraction that is followed by stacking multiple classification models for detecting malignant and benign lesions. The novelty of the proposed method is due to involvement of a stacking ensemble that is trained using the features extracted by CNN models. The ensemble model contains four classifiers (SVM, KNN, ANN, and RF) and performances of features extracted by ResNet50, VGG16 and Xception models are compared for suitability to the proposed model.

3 MATERIALS AND METHOD
3.1 The Dataset

Dermoscopy images provided by the International Skin Imaging Collaboration (ISIC) are used in this work. The dataset is a collection of 1800 benign and 1497 malignant images all of which have dimensions of 224×224 pixels [25]. It is randomly split into training and test sets with 70% and 30%, respectively. Some sample images of the dataset are given Fig. 1.
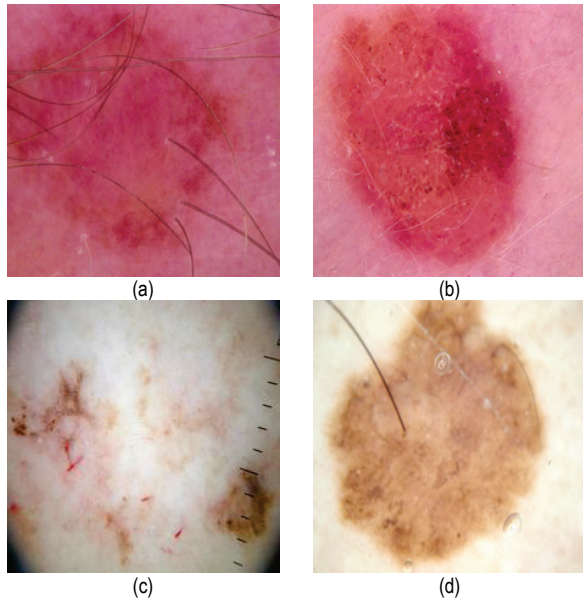
Figure 1 Benign (a, b) and malignant (c, d) samples from the dataset

3.2 Stacked Cross Validation with Deep Features

It is possible to divide the proposed SCV-DF method into three main levels of processing. The first level involves feature extraction from the images. The melanoma images in the dataset are structured data and the feature extraction from these images is an important task for an accurate classification. Deep learning methods allow for automatic extraction of the features from the images and there are state-of-the art CNN architectures to perform this task. Therefore, deep features are extracted from the images using three different CNN architectures. Features from these architectures are analyzed separately to understand the appropriateness of them. These features are used in the second level for training on the base models in SCV-DF method. The prediction outputs of the base models are stacked to be used in the third level. Together with the actual target values, the stack of the predictions is fed into a meta-model in the level three. The output of the meta-model is the final prediction that is used to evaluate the performance of the method.

3.2.1 Feature Extraction

Feature extraction is related to the first processing level of the proposed method. CNN architectures generally have two major parts. First one is feature extraction part in which there are convolutional layers together with activation, regularization and pooling operations. The second one is the classification part that contains several fully connected layers followed by a final decision layer.

In this work, the features used by the SCV-DF algorithm are extracted through convolutional layers of three different deep learning architectures that are ResNet50 [26], VGG16 [27] and Xception [28]. ResNet50 is a CNN architecture with a depth of 50 layers. The associated kernel size is 7×7 and 64 different kernels with a stride 2 is used for each layer.

VGG16 uses only 3×3 convolutions stacked on top of each other where max pooling is used to reduce the volume size. It consists of 16 weight layers in which 13 convolutional layers and three fully connected layers. Xception architecture contains linear stack of depths separable convolutions together with residual connections to eliminate the risk of vanishing gradient. Xception model has 36 convolutional layers in its feature extraction base.

Training of the deep learning models were performed with stochastic gradient descent optimizer and binary cross entropy was used as the loss function. The learning rate and the momentum parameters of the optimizer were set as 0.01 and 0.9, respectively.

3.2.2 Stacked Cross Validation Algorithm

The second and third levels of the SCV-DF method are involved in stacked CV algorithm that contains several base models and a meta-model that performs the final decision. In this algorithm, the base models are typically different and they are all trained on the same training set. The base models used in this work are SVM, KNN, ANN, and RF. Together with the expected outputs, the predictions made by the base models are fed into the meta-model to learn a relationship between the inputs and outputs. The most common meta-model is a logistic regression function which is used in this work as well. In order to include the cross-validation property, the dataset is divided into k folds and one fold is held out for testing, and the remaining k–1 folds are used for training for the base models. This procedure is repeated until all the folds are used as a test set. The prediction outputs of these models on the test sets of k folds are stacked to be used in the third level. In this algorithm, k is chosen as five. which is found through the experimentation to estimate with low bias a modest variance. All three levels of the proposed method are illustrated in Fig. 2.

3.3 Performance Evaluation Metrics

For the performance evaluation of this work, accuracy, sensitivity, *F1-score*, and AUC are computed. Correct calculation of these measures requires careful definition of the terms true positive (*TP*), true negative (*TN*), false positive (*FP*) and false negative (*FN*). Since detection of malignant type is the main concern, correct classification of an image containing malignant tissue is considered as a *TP* prediction. The other terms are defined accordingly and they are summarized in the confusion matrix given in Tab. 1.

Table 1 The Confusion Matrix

		Actual Class	
		Benign	Malignant
Predicted Class	Benign	*TN*	*FN*
	Malignant	*FP*	*TP*

The receiver operating characteristic curve (ROC curve) corresponds to the performance of the proposed model at all classification thresholds. It is the plot for true positive rate versus false positive rate. The area under the ROC curve

(AUC) provides the aggregate measure of all possible classification thresholds. The other three performance metrics are accuracy, sensitivity, and *F1-score* that are calculated using Eqs. (1), (2), and (3), respectively. *Accuracy* is a measure for rate of correct predictions among all samples in the test set. *Sensitivity*, also known as recall, is the ratio of correctly predicted malignant samples to the all actually malignant samples. On the other hand, *F1-score* is an important measure for imbalanced dataset.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{2}$$

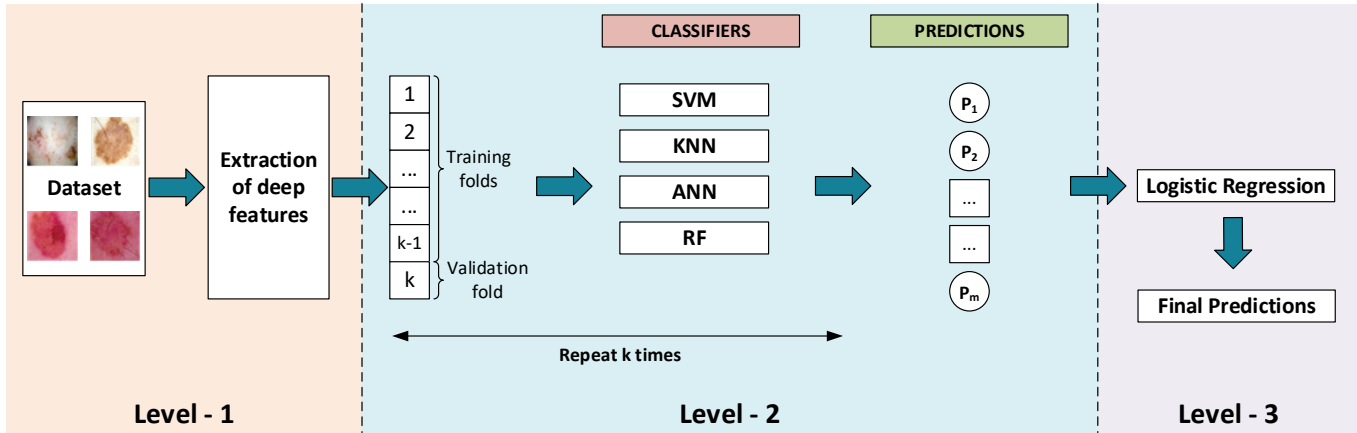$$F1\text{-}score = \frac{2TP}{2TP + FP + FN}. \tag{3}$$

Figure 2 The processing pipeline of the proposed SCV-DF method

4 RESULTS AND DISCUSSION
4.1 Evaluation of the Proposed Method

As indicated earlier, the first layer of SCV-DF involves utilization of deep learning methods for feature extraction from the images. In order to compare performances of different CNN architectures, three baseline models, namely, ResNet50, Xception, and VGG16, were used. Some of the feature maps generated with these models are given in Figs. 3 and 4 for samples of benign and malignant tissues, respectively. Since all these CNN models have high number of layers, each of which contains many filters, only some outputs of the filters are included in these figures. In order to illustrate how the frequency components of each layer change at different depths of processing, feature maps obtained at three different layers (2nd, 9th, and 13th) are provided in the corresponding columns. As can be seen from the figures that the feature maps belonging to initial layers resemble the input image more than those generated at the higher layers. In other words, as the image is processed at the deeper layers of the network, the meaningful features for the human eye are replaced by the features that are important for the classification model. Furthermore, the sizes of the feature maps are reduced by max-pooling operations at various layers. As a result, the feature maps obtained at higher layers have lower resolution and this situation is observable in Figs. 3 and 4, as well. Despite their low resolution, the number of such feature maps is typically high in CNN models and they are flattened into one single high dimensional feature vector which is the output of the first layer of processing the proposed method.
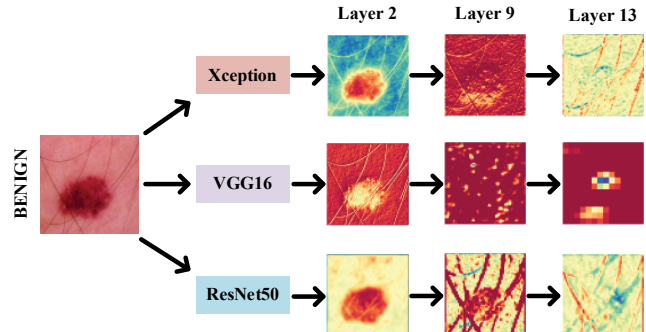

Figure 3 Feature maps generated by CNN architectures for a benign sample
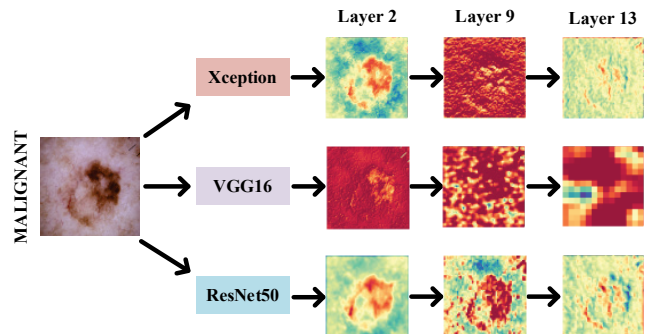

Figure 4 Feature maps generated by CNN architectures for a malignant sample

Next in the second layer, the extracted features were used for training of a stacked CV model for each of these architectures. To underline the appropriateness of stacked CV for this task, six different single benchmark classifiers were trained using these extracted features as well. These classifiers are SVM, KNN, ANN, RF, logistic regression, and

AdaBoost. The first five of them are the individual classifiers that are used in the second and third levels of the stacked CV method. Like stacked CV, AdaBoost is an ensemble learning method in which outputs of several weak learners are combined. The reason for selecting these benchmark methods is to underline the usefulness of stacked CV instead of using them individually. The details about the model parameters of single benchmark classifiers are provided in Tab. 2. When selecting the model parameters, their default values were used initially. Then the effects of slightly changing these values were observed to determine them.

Table 2 Properties and parameters of the benchmark models

Method	Properties
SVM	Kernel: Gaussian, C = 1, gamma = 0.001
KNN	K=5
ANN	# hidden units = 100, learning rate = 0.0001
RF	# trees = 100, split quality measure: Gini index
Logistic Regression	Penalization: L2-norm
AdaBoost	Base learner: Decision tree, # weak learners: 50

As can be seen from Tabs. 3-6 that the highest values for all of the four performance metrics are achieved by using the features extracted via Xception architecture in SCV-DF method. When compared to performances of individual benchmark models, SCV-DF with ResNet50 and VGG16 features improves the classification performance in general. In particular, *F1-score* and *AUC* values of SCV-DF with ResNet50 outperform all the individual models. On the other hand, the ensemble model, AdaBoost, is capable of achieving higher scores than SCV-DF model with ResNet50 features. However, this situation is not observable for all types of features. For example, with VGG16 features, the AdaBoost model has relatively lower accuracy, *F1-score*, and sensitivity values than those obtained via ResNet50 features. On the other hand, VGG16 features achieve higher *AUC* values with all benchmark models than ResNet50 features.

Table 3 Accuracy Values Obtained with SCV-DF and the Other Methods

	ResNet50	Xception	VGG16
SCV-DF	0.816	0.909	0.865
SVM	0.818	0.867	0.867
KNN	0.812	0.864	0.834
ANN	0.780	0.801	0.831
RF	0.787	0.803	0.840
Logistic Regression	0.806	0.863	0.875
AdaBoost	0.821	0.837	0.799

Table 4 F1-Score Values Obtained with SCV-DF and the Other Methods

	ResNet50	Xception	VGG16
SCV-DF	0.788	0.896	0.842
SVM	0.787	0.838	0.835
KNN	0.778	0.834	0.803
ANN	0.785	0.720	0.753
RF	0.735	0.739	0.805
Logistic Regression	0.762	0.837	0.847
AdaBoost	0.792	0.739	0.766

Even though it is not possible to tell a best deep feature extractor architecture that outperforms the other two in all metrics and for all classifiers, the highest scores are obtained by Xception architecture. As a result, it may be concluded that the SCV-DF method can improve the detection

performance when it is used together with appropriate CNN architecture for feature extraction. Bar graphs for the accuracy values are given in Fig. 5 for visual comparison.

Table 5 Sensitivity Values Obtained with SCV-DF and the Other Methods

	ResNet50	Xception	VGG16
SCV-DF	0.821	0.886	0.804
SVM	0.816	0.834	0.810
KNN	0.736	0.827	0.799
ANN	0.740	0.783	0.776
RF	0.715	0.678	0.798
Logistic Regression	0.752	0.853	0.844
AdaBoost	0.825	0.798	0.794

Table 6 AUC Values Obtained with SCV-DF and the Other Methods

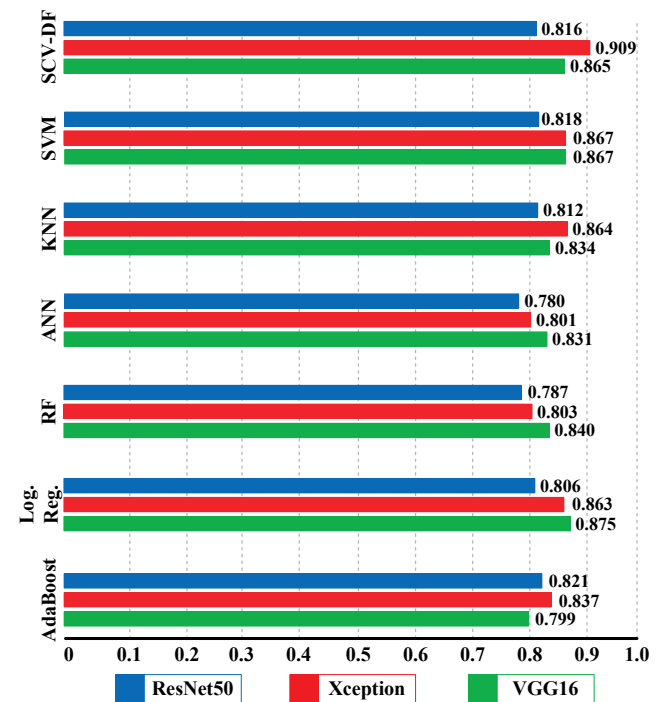|	ResNet50	Xception	VGG16
SCV-DF	0.818	0.917	0.843
SVM	0.817	0.862	0.859
KNN	0.762	0.861	0.834
ANN	0.781	0.820	0.831
RF	0.777	0.784	0.834
Logistic Regression	0.798	0.862	0.870
AdaBoost	0.822	0.831	0.834

Figure 5 Percentage of accuracy values for visual comparison of the methods

The burden of false negative classifications is higher than false positives in this work because misclassification of actual unhealthy images may result in late diagnosis of the disease. As a result, it may cause catastrophic consequences for the patient. Sensitivity is measured is related to the rate of false negatives hence; extra emphasis needs to be put on this measure. Obviously, SCV-DF with Xception features achieves the highest sensitivity value as 0.886. It is also notable that SCV-DF with the other two deep learning based feature extraction methods can improve the sensitivity as well.

The AUC values are obtained by calculating the area under the ROC curves for each of the classifiers respectively.

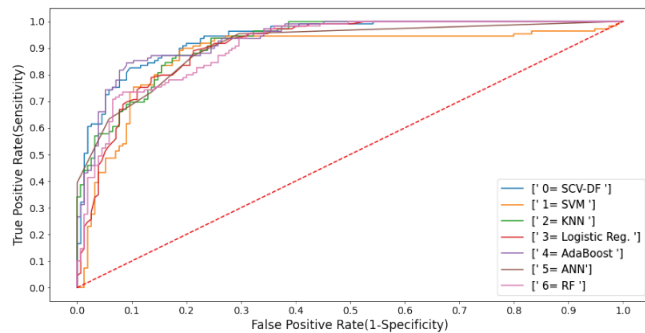The graphical representation of the related ROC curves is shown in Fig. 6.

Figure 6 ROC curves of the classifiers

4.2 Comparison with the Existing Studies

In the literature, there are other studies in which the images of skin lesions are classified into different categories. In this section, the results obtained by SCV-DF are compared by those reported in five of the recent works where malignant and healthy lesions are detected.

In the majority of the other studies, the researchers generally utilize deep learning methods as they allow for automatic extraction of features. The obtained accuracy and AUC values, together with the utilized methods in these studies, are summarized in Tab. 7.

Both the accuracy and AUC scores achieved by the proposed method is higher than the other studies in the literature which supports the effectiveness of SCV-DF algorithm for detection of malignant skin tissue from dermoscopic images.

Table 7 Summary of the Existing Studies in the Literature

Reference No	Year	Method	*Accuracy*	*AUC*
[29]	2017	Sequential CNN	72.1%	0.94
[30]	2018	InceptionV4 + SVM	89%	0.773
		InceptionV4 + NN	89%	0.892
[31]	2018	ResNet50	72.1%	0.888
[32]	2021	AlexNet	81.8%	
		Sequential CNN	82.01%	-
		ResNet50	90.78%	
[33]	2021	VGG16	80%	-
		ResNet50	87%	-
Proposed method	2021	SCV-DF	90.9%	0.964

5 CONCLUSION

SCV-DF, a hybrid method for classification of dermoscopy images, is proposed in this work. The method includes extraction of deep features from the images using the convolutional layers of three deep learning architectures, which are ResNet50, Xception, and VGG16. These features are fed into a stacked-CV step where four different baseline classifiers are trained and their prediction results are merged. Together with the actual labels, these prediction results are then used as a training set for a meta-classifier. The baseline classifiers are SVM, KNN, ANN, and RF, and the meta-classifier is a logistic regression model.

The method is developed and tested using a dataset containing 1800 benign and 1497 malignant images. The performance of the proposed method is compared with the cases in which the baseline classifiers and the meta-classifier are trained individually on the deep features. Furthermore, as an alternative ensemble learning method, the performance of the AdaBoost method is evaluated to compare it with the stacked CV. According to the results, it has been shown that SCV-DF outperforms the benchmark models when deep features are extracted using the Xception network. Furthermore, all four of the calculated performance metrics for SCV-DF is higher than AdaBoost when VGG16 and Xception networks are used for feature extraction. In addition, the accuracy and *AUC* values of the proposed method were compared with the reported results in the relevant literature. It has been shown that SCV-DF model outperforms the other deep learning based methods. Therefore, it may be concluded that the proposed SCV-DF method is suitable for detection of malignant skin cancer lesions. In more general terms, this study has revealed that ensemble methods can increase the detection accuracy for this specific problem particularly when the features are extracted with the appropriate CNN architecture.

As a future work, sub-classes of the malignant and benign tissues is expected to be detected with an improved version of the method. In addition, other state-of-the-art CNN architectures are going to be included in the experiment to understand their suitability for this task.

6 REFERENCES

[1] American Cancer Society. (2020). Risk Factors for Melanoma Skin Cancer. www.cancer.org/cancer/melanoma-skin-cancer/causes-risks-prevention/risk-factors.html. Accessed 15 April 2021.

[2] Chaturvedi, S. S., Gupta, K., & Prasad, P. S. (2020). Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using MobileNet. Paper presented at *the International Conference on Advanced Machine Learning Technologies and Applications*. https://doi.org/10.1007/978-981-15-3383-9_15

[3] American Cancer Society. (2019). Cancer Facts & Figures. www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf. Accessed 15 April 2021.

[4] Zghal, N. S. & Derbel, N. (2020) Melanoma Skin Cancer Detection based on Image Processing. *Current Medical Imaging, 16*(1), 50-58. https://doi.org/10.2174/1573405614666180911120546

[5] Polat, K. & Koc, K. O. (2020). Detection of skin diseases from dermoscopy image using the combination of convolutional neural network and one-versus-all. *Journal of Artificial Intelligence and Systems, 2*(1), 80-97. https://doi.org/10.33969/AIS.2020.21006

[6] Wei, L., Ding, K., & Hu, H. (2020). Automatic Skin Cancer Detection in Dermoscopy Images based on Ensemble Lightweight Deep Learning Network. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2997710

[7] Ech-Cherif, A., Misbhauddin, M., & Ech-Cherif, M. (2019). Deep Neural Network based mobile dermoscopy application for triaging skin cancer detection. Paper presented at *the 2nd International Conference on Computer Applications & Information Security (ICCAIS'19)*.

https://doi.org/10.1109/CAIS.2019.8769517

[8] Yélamos, O., Braun, R. P., Liopyris, K., et al. (2019). Usefulness of dermoscopy to improve the clinical and histopathologic diagnosis of skin cancers. *Journal of the American Academy of Dermatology, 80*(2), 365-377. https://doi.org/10.1016/j.jaad.2018.07.072

[9] Wolner, Z. J., Yélamos, O., Liopyris, K., Rogers, T., Marchetti, M. A., & Marghoob, A. A. (2017). Enhancing skin cancer diagnosis with dermoscopy. *Dermatologic clinics, 35*(4), 417-437. https://doi.org/10.1016/j.det.2017.06.003

[10] Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A. (2019). Cancer diagnosis using deep learning: a bibliographic review. *Cancers, 11*(9), 1235. https://doi.org/10.3390/cancers11091235

[11] Rajasekhar, K. S. & Babu, T. R. (2020). Analysis and Classification of Dermoscopic Images Using Spectral Graph Wavelet Transform. *Periodica Polytechnica Electrical Engineering and Computer Science*. https://doi.org/10.3311/PPee.15048

[12] Arasi, M. A., El-Horbaty, E.-S. M., & El-Sayed, A. (2018). Classification of Dermoscopy Images Using Naïve Bayesian and Decision Tree Techniques. Paper presented at *the 1st Annual International Conference on Information and Sciences (AiCIS'18)*. https://doi.org/10.1109/AiCIS.2018.00015

[13] Murugan, A., Nair, S. A. H., & Kumar, K. S. (2019) Detection of skin cancer using SVM, Random Forest and kNN classifiers. *Journal of medical systems, 43*(8), 269. https://doi.org/10.1007/s10916-019-1400-8

[14] Kavitha, J. & Suruliandi, A. (2016). Texture and color feature extraction for classification of melanoma using SVM. Paper presented at *the International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)*. https://doi.org/10.1109/ICCTIDE.2016.7725347

[15] Ezat, W., Dessouky, M., & Ismail, N. (2020). Multi-class Image Classification Using Deep Learning Algorithm. Paper presented at *the Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1447/1/012021

[16] Jianu, S. R. S., Ichim, L., & Popescu, D. (2019). Automatic diagnosis of skin cancer using neural networks. Paper presented at the *11th International Symposium on Advanced Topics in Electrical Engineering (ATEE'19)*. https://doi.org/10.1109/ATEE.2019.8724938

[17] Iliadis, L. S. & Jayne, C. (2020). Emerging applications of Deep Learning and Spiking ANN. *Neural Computing and Applications, 32*(23), 17119-17124. https://doi.org/10.1007/s00521-020-05443-z

[18] Lynn, N. C. & Kyu, Z. M. (2017, 18-20 Dec. 2017). Segmentation and Classification of Skin Cancer Melanoma from Skin Lesion Images. Paper presented at *the 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'17)*. https://doi.org/10.1109/PDCAT.2017.00028

[19] Goyal, M., Oakley, A., Bansal, P., Dancey, D., & Yap, M. H. (2019). Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. *IEEE Access, 8*, 4171-4181. https://doi.org/10.1109/ACCESS.2019.2960504

[20] Seeja, R. & Suresh, A. (2019). Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (svm). *Asian Pacific Journal of Cancer Prevention: APJCP, 20*(5), 1555. https://doi.org/10.31557/APJCP.2019.20.5.1555

[21] Hekler, A., Utikal, J. S., Enk, A. H., et al. (2019). Superior skin cancer classification by the combination of human and artificial intelligence. *European Journal of Cancer, 120*, 114-121. https://doi.org/10.1016/j.ejca.2019.07.019

[22] Kassani, S. H. & Kassani, P. H. (2019). A comparative study of deep learning architectures on melanoma detection. *Tissue and Cell, 58*, 76-83. https://doi.org/10.1016/j.tice.2019.04.009

[23] Codella, N. C., Nguyen, Q.-B., Pankanti, S., et al. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development, 61*(4/5), 5: 1-5: 15. https://doi.org/10.1147/JRD.2017.2708299

[24] Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe, Y., & Hamamoto, R. (2020). The development of a skin cancer classification system for pigmented skin lesions using deep learning. *Biomolecules, 10*(8), 1123. https://doi.org/10.3390/biom10081123

[25] Ruthra, V. & Sumathy, P. (2019). Color and Texture based Feature Extraction for Classifying Skin Cancer using Support Vector Machine and Convolutional Neural Network. *International Research Journal of Engineering and Technology, 6*(9), 502-507.

[26] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. Paper presented at *the European conference on computer vision*. https://doi.org/10.1007/978-3-319-46493-0_38

[27] Simonyan, K. & Zisserman, A. (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[28] Chollet, F. (2017, 21-26 July 2017). Xception: Deep Learning with Depthwise Separable Convolutions. Paper presented at *the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2017.195

[29] Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118. https://doi.org/10.1038/nature21056

[30] Pham, T.-C., Luong, C.-M., Visani, M., & Hoang, V.-D. (2018). Deep CNN and data augmentation for skin lesion classification. Paper presented at *the Asian Conference on Intelligent Information and Database Systems*. https://doi.org/10.1007/978-3-319-75420-8_54

[31] Yap, J., Yolland, W., & Tschandl, P. (2018). Multimodal skin lesion classification using deep learning. *Experimental dermatology, 27*(11), 1261-1267. https://doi.org/10.1111/exd.13777

[32] Poornima, A., Devi, M. S., Sumithra, M., et al. (2021). Epoch interrogation for skin cancer detection using convolutional neural network models. Paper presented at *the IOP Conference Series: Materials Science and Engineering*. https://doi.org/10.1088/1757-899X/1074/1/012025

[33] Manasa, K. & Murthy, D. G. V. (2021). Skin Cancer Detection Using VGG-16. *European Journal of Molecular & Clinical Medicine, 8*(1), 1419-1426.

Authors' contacts:

Ahmed Al-Karawi, MSc
Çukurova University,
Faculty of Engineering,
Electrical & Electronics Engineering Department,
01330, Balcalı, Sarıçam, Adana, Turkey
+903223386868, alzeerahmed1993@gmail.com

Ercan Avşar, PhD
(Corresponding author)
Dokuz Eylül University, Computer Engineering Department,
Tınaztepe Campus, Buca 35160, Izmir, Turkey
+902323017401, ercan.avsar@deu.edu.tr