Highlights

A novel soft attention-based multi-modal deep learning framework for multi-label skin lesion classification

Asli Nur Omeroglu,Hussein M.A.Mohammed,Emin Argun Oral,Serdar Aydin

- Soft attention-based multi-modal network is proposed for multi-label classification
- The proposed network applied for multi-label skin lesion classification problem
- Multiple branches are efficiently integrated to learn the complementary features

A novel soft attention-based multi-modal deep learning framework for multi-label skin lesion classification

Asli Nur Omeroglu[a,*], Hussein M.A.Mohammed[a], Emin Argun Oral[a,b,**] and Serdar Aydin[a,b]

[a]Ataturk University, Department of Electrical and Electronics Engineering, Yakutiye, Erzurum 25240, Turkey,
[b]Ataturk University, High Performance Computing Application and Research Center, Yakutiye, Erzurum, 25240, Turkey,

ARTICLE INFO

Keywords:
Multi-modal fusion
Multi-label classification
Skin lesion diagnosis
Seven-points checklist criteria

Abstract

Skin cancer is one of the fatal cancers worldwide. Early detection of this disease can significantly increase the survival rate. In this study, a multi-modal and soft attention based hybrid deep learning model is proposed for automated and accurate multi-label skin lesion classification. The proposed network includes a multi-branch structure that integrates feature maps from different modalities in a hybrid manner. These branches enable to extract features from multiple modalities separately and learn complex combinations between them. These branches include a modified Xception architecture, a new feature extraction method, as well as soft attention module that is proposed to make the network focus on discriminative parts of skin lesions. The final diagnosis is obtained by the fusion of the predictions from three branches. The proposed framework was evaluated on the publicly available seven-point criteria evaluation dataset, a well-established multi-modality multi-label skin disease dataset. It achieved an average accuracy of 83.04% for multi-label skin lesion classification. It is more accurate than the state-of-the-art methods and improves the average accuracy by more than 2.14% on the test set.

1. Introduction

Skin cancer is one of the common types of cancer, and its incidence rates have been steadily increasing (Siegel et al., 2019). Melanoma, a fatal form of skin cancer, has a 10-year survival rate of below 39% if detected at advanced stage, while the early detection of melanoma can increase the 10-years survival rate over 93% (Balch et al., 2009). Therefore, early diagnosis of melanoma is important to improve survival rates. Artificial intelligence (AI), especially machine learning (ML), on the other hand, has grown rapidly in recent years and has been used in different applications such as control systems, materials science, and healthcare (Goyal et al., 2020, Stojanovic and Nedic, 2016, Tao et al., 2020, Versaci et al., 2022, Zhang et al., 2021, Zhou et al., 2022). Recent advances in the application of machine learning techniques have led to a great progress in computer-aided implementations (Antonelli et al., 2014, Roh et al., 2014). Utilizing machine learning techniques can automate the screening and early detection of skin lesion classification in case of facility and specialist deficiencies, especially in rural areas (Adegun and Viriri, 2021).

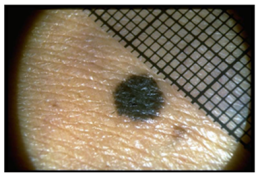
Skin cancer is generally diagnosed by dermatologists' visual examination. They use clinical and dermoscopy imaging for this examination. Clinical images are acquired using a digital camera. They provide highly detailed information about the important global features of lesions such as elevation, geometry, borders, and color, although they cannot depict the subsurface details of lesions. They are less standardized as the lighting and the field of vision may differ between images of different sources. Also, they sometimes may include some visually blocking components such as tattoos or body piercings. Dermoscopy images, on the other hand, are acquired by using specific skin imaging techniques such as liquid immersion or cross-polarized lighting to cut down surface reflection, and they offer a magnified (usually 10x) microscopic image of the lesion. They present detailed morphological and visual properties of lesions such as pigmentation distribution, dot shape, and streak presentation. Additionally, they display subsurface features like dots, globules, and vascularity. Dermoscopy images are more comprehensive and offer a thorough examination of the lesion. They have been, for example, shown to increase the accuracy rate of melanoma diagnosis (Carli et al., 2004, Vestergaard et al., 2008). Combining dermoscopy and clinical imaging is currently the standard procedure clinical practice to consider macroscopic and microscopic aspects (Puig et al., 2007).

Abbreviations: ASB, Absent; ATY, Atypical; AUC, Area Under the Curve; AVG, Average Accuracy; BCC, Basal Cell Carcinoma; BWV, Black Whitish Veil; CAD, Computer Aided Diagnosis; CC, Classifier Chain; CIM, Clinical Image Modality; CNN, Convolutional Neural Network; DAG, Dots and Globules; DCIHMM, late fusion of DIM, CIM and HMM; DCIM, Dermoscopy and Clinical Image Modality Fusion; DIAG, Diagnosis; DIM, Dermoscopy Image Modality; HMM, Hybrid Multi-Modality; IRG, Irregular; MA, Mutual Attention; MEL, Melanoma; MISC, Miscellaneous; NEV, Nevus; PIG, Pigment; PN, Pigment Network; PRE, Precision; PRS, Present; REG,Regular; ROC, Receiver Operating Characteristic Curve; ROI, Region of Interest; RS, Regression Structures; SEN, Sensitivity; SK, Seborrheic Keratosis; SPC, Seven-Point Checklist; SPE, Specificity; STR, Streak; SVM, Support Vector Machine; TYP, Typical; VS, Vascular Structures

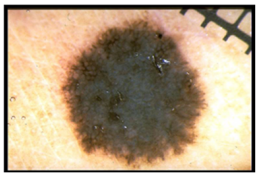
*Corresponding author
**Principal Corresponding author.
asli.omeroglu@atauni.edu.tr (A.N. Omeroglu); halameri080@gmail.com (H. M.A.Mohammed); eminoral@atauni.edu.tr (E.A. Oral); serdar@atauni.edu.tr (S. Aydin)
ORCID(s):

Conventional diagnosis of melanoma has relied on rule-based diagnostic strategies such as ABCD rules (Nachbar et al., 1994), 7-point checklist (Argenziano et al., 2011), and 3-point checklist (Soyer et al., 2004). Dermatologists have commonly utilized these strategies to enable the visual identification of malignancy related characteristics and simplify diagnostic processes.

As shown in Fig. 1, the seven-point checklist (SPC), one of the main approaches, includes seven criteria for the simultaneous assessment of different visual features of lesion, and each criterion is assigned a score according to its importance. In this approach, major criteria are assigned 2 points while the minor ones are assigned 0 point. The lesion is diagnosed as melanoma if the total score exceeds the threshold of 3 points.



Clinical Image



Dermoscopy Image

Seven Checklist Criteria	Category	Score
Blue Whitish Veil (BWV)	Present	2
Dots and Globules (DaG)	Irregular	1
Pigmentation (PIG)	Absent	0
Pigment Network (PN)	Atypical	2
Regression Structure (RS)	Absent	0
Streaks (STR)	Regular	0
Vascular Structures (VS)	Absent	0
Diagnosis (DIAG)	Melanoma	Total:5

Patient's Meta-Data	
Diagnostic Difficulty	High
Location	Lower limbs
Sex	Female
Management	Excision
Elevation	Flat

Fig. 1. An Example of a patient's data, including the clinical image and dermoscopy image along with the meta-data. The corresponding seven-points checklist criteria, its scores and the diagnosis label are also shown. (Kawahara et al., 2018).

Assessment of clinical and dermoscopy image with human vision alone is a highly subjective and complex task in the SPC approach. Since many labels require to be evaluated simultaneously and may differ for each criterion. Thus, recently researchers have been directed on developing Computer Aided Diagnosis (CAD) systems to assist in dermatologists' clinical diagnosis (Cheng et al., 2016, Pathan et al., 2018). CAD systems can both support clinicians during their daily clinical routine and enable fast as well as cheap access to improved diagnostic accuracy, even outside of the clinics by the use of mobile device based applications. (de Carvalho et al., 2019).

1.1. Related work

This section reviews methods applied for the classification of skin lesions in the literature. They fall into three categories. In the first, machine learning and deep learning-based methods using single-modal data are examined. In the second, deep learning methods utilizing multi-modal data are presented. And finally, multi-label and multi-modal based deep learning approaches for skin lesions classification are summarized in the last category.

1.1.1. Skin lesion classification based on single-modality data

Early automated skin lesion classification approaches, generally based on dermoscopy images, can be divided into traditional methods and Convolutional Neural Network (CNN) based methods. In traditional methods, many handcrafted features related to the shape, texture and color of the lesion are first extracted and then powerful classifiers such as Support Vector Machines (SVMs), AdaBoost, and Random forest are applied (Arora et al., 2020, Dhivyaa et al., 2020, Fisher et al., 2019, Mporas et al., 2020). These handcrafted feature based approaches, however, are significantly rely on image preprocessing and expert specialist knowledge to compute visual descriptors for each category. Furthermore, these approaches are not robust to dermoscopy images, as dermoscopy images are captured with varying imaging equipment and lighting conditions.

Convolutional neural network based approaches and their variants, on the other hand, have been proved to be effective across a wide range of image related problems and have been applied to skin lesion diagnosis due to the flexibility of CNN structures and their outstanding performance (He et al., 2016, Simonyan and Zisserman, 2014). They have been demonstrated to achieve levels of accuracy very similar to those of dermatologists in detecting several different types of skin lesions (Brinker et al., 2019, Haenssle et al., 2018). Recently published CNN based skin lesion classification methods (Kawahara et al., 2016, Menegola et al., 2017, Tang et al., 2019, 2020, Yu et al., 2016, 2018, Zhang et al., 2019) primarily focus on transfer-learning models that fuse deep and handcrafted features, segmentation-classification models, and self-attention modules. (Kawahara et al., 2016) and (Benyahia et al., 2022) proposed a CNN model which is finetuned with the pre-trained weight on large datasets, and (Tschandl et al., 2019) proposed a similar approach that uses pretrained three deep learning models with hybrid approach. These methods outperformed the traditional methods. (Sharafudeen et al., 2022, Sharma et al., 2022) and (Yu et al., 2018) adopted the combination of handcrafted and deep feature vectors that are derived from CNNs, and the results of the proposed models show higher classification accuracy than CNN models alone. (Hasan et al., 2022), (Khan et al., 2021c) and (Khan et al., 2021b) used DSNet, Mask RCNN and CNN, respectively, to segment the region of interest (ROI) from dermoscopy images. Then cropped images were classified based on ROI for lesion detection. Several studies have also demonstrated that segmentation-classification algorithms outperform single-classification methods in terms of accuracy (Tang et al., 2019, Zhang et al., 2019). (Ding et al., 2021, Gessert et al., 2019, Wu et al., 2020) and (Abdelhalim et al., 2021) proposed the attention module instead of the segmentation model to prioritize the skin lesion area. This attention module improves classification performance without extra segmentation labels.

Although these methods have good performance, they rely on single image classification, and they do not reflect clinical practice. In fact, dermatologists make decisions based both on visual inspections of images from different sources and on the integration of information from patients' meta-data such as family medical history, ethnicity, social habits, and sun exposure.

1.1.2. Skin lesion classification based on multi-modality data

Classification of skin lesions using different source of skin lesion images and patient demographics – also known as patient data – can be considered a multi-modal fusion procedure. There are few studies for multi-modal skin lesion classification. (Ge et al., 2017) demonstrated the effectiveness of multi-modal learning using clinical and dermoscopy images by comparing the proposed triplet networks. (Yap et al., 2018) proposed two ResNet-50 models to extract feature vectors from clinical and dermoscopy images, and these feature vectors were then combined with meta-data (e.g., age, sex). Similarly, (Liu et al., 2020) and (Chen et al., 2022) proposed a skin lesion diagnostic system that combines skin images and clinical medical history to effectively improve diagnosis. These methods are generally designed for fusing representations of each modality at the end of a CNN model.

There are others that gradually fuse features of different modalities on all levels of CNN architecture. (Li et al., 2020) proposed a multiplication method. This method used clinical information features to recalibrate the weight of image features and performed better than the cascade data fusion approach. (Wang et al., 2021a) proposed a method, three-branch IM-CNN structure, that gradually combines meta-data, features from segmented skin lesions, and skin lesion images, respectively. (Cai et al., 2022) proposed the Mutual Attention (MA) block to better combine image features and meta-data features.

An alternative to such multi-modality approaches for skin lesion classification problem is to design a multi-tasking architecture that consider multiple criteria such as 7-point checklist to obtain more interpretable models.

1.1.3. Multi-label skin lesion classification based on multi-modality data

Multi-modal methods using 7-point checklist criteria for skin lesion classification are less common. (Kawahara et al., 2018) first introduced a multi-modal architecture that used a pretrained Inception network using the patients' dermoscopic images, clinical images, and meta-data to predict malignancy based on the 7-point checklist criteria. (Fu et al., 2022) proposed a graph based model that consists of inter-category and inter-modality network (GIIN) modules using dermoscopy and clinical images. (Wang et al., 2021b) proposed a method combining a multi-label deep feature extractor and a clinically constrained classifier chain (CC). In this method, deep convolutional neural networks combine dermoscopy images and clinical images, and the constrained classifier chain encodes relevance between different criteria. (Bi et al., 2020) designed a hyperbranched fusion block that combines feature maps from dermoscopy and clinical branches at different scales to classify multiple tags in multi-modality skin lesion images. (Tang et al., 2022) proposed a two stage fusion algorithm that gradually integrates the information from three modality data (dermoscopy image, clinical image, and meta-data). (Wang et al., 2022) proposed a self attention based image reconstruction approach by using two modalities for multi-modal multi-label skin classification.

1.2. Contribution of the proposed study

The attention strategy, that can be trained to focus on critical regions, improves network performance in image classification tasks (Bai et al., 2021, Yan et al., 2019). Some researchers have applied various attention mechanisms in analyzing medical images (Hassanzadeh et al., 2021, Sitaula and Hossain, 2021). Although these models have promising results for the skin lesion classification task, most of them focus on a single lesion type and different modalities are not considered (Hasan et al., 2021, Khan et al., 2021a, Liang et al., 2019). There is still room for an improvement in skin lesion classification, and better performances can be achieved by using a better multi-modal fusion methodology. Motivated by this approach, a hybrid soft attention-based multi-modal deep learning model is proposed in this paper for multi-label skin lesion classification. Here, a multi-branch architecture consisting of three parallel subnets is designed to effectively learn the complementary features of different modalities. A modified Xception network with soft attention module is proposed as the main source of feature extraction for three separate branches. This soft attention module is used to focus on relevant lesion areas. Multiple inputs are fused in a hybrid manner to provide maximum complementary information transfer between different modalities. Finally, a late fusion scheme that efficiently combines the outputs of three branches is proposed to obtain multi-label classification results.

The contributions of this paper are summarized as follows:

- A novel multi-modal and multiple attention learning model, based on modified Xception architecture, is proposed for multi-label skin lesion classification.
- The proposed framework is capable of integrating complementary global features of different modalities through the use of three branches in a hybrid manner backed by soft attention mechanism.
- To our knowledge, this is the first study that exploits a multi-modal approach based on hybrid soft attention mechanism for multi-label skin lesion classification.
- Comprehensive experiments have been conducted on publicly available *7-point criteria evaluation* dataset to verify the effectiveness of the proposed framework.

- Comparison of the obtained results with the literature shows that the proposed method achieves state-of-the-art performance with an average accuracy rate of 83.04% for multi-label skin lesion classification problem.

2. Proposed method

In this section, the proposed automated multi-label skin lesion classification framework is introduced. It is a hybrid attention deep learning model, based on multi-modal data, as shown in Fig. 2. These are dermoscopy imaging modality branch, clinical imaging modality branch, and hybrid-meta branch. The dermoscopy and clinical imaging modality branches use dermoscopy and clinical images as their input for feature extraction, respectively. Extracted features of these two modalities are concatenated and then applied to the hybrid-meta branch as its input. Features of three sources, namely clinical image, dermoscopy image, and their concatenation, are further processed by separate soft attention modules. Besides these, the feature set obtained from the third modality, namely meta-data, is simultaneously concatenated with that of the hybrid-meta branch. Each branch is terminated by a softmax layer, as shown in Fig. 2. Finally, the outputs of all three branches are then fused at the decision level for multi-label skin lesion classification. Details of the proposed method are explained below.

2.1. Dermoscopy and clinical imaging modality branches

The illustration of dermoscopy and clinical imaging modality branches are shown in Fig. 2. The dermoscopy and clinical imaging branches utilizes modified Xception framework as the backbone structure for feature extraction.

Xception has been used for many tasks related to skin lesions, such as automated classification and segmentation (Adegun and Viriri, 2021, Calderón et al., 2021, Panthakkan et al., 2022). Because of its contribution to the performance improvement, it was utilized in the proposed study as also employed by other research groups. Xception network is capable of producing an increase in accuracy as well as a significant decrease in computational cost by using depthwise separable convolutions and residual connections (Chollet, 2017). It is based on the Inception architecture that its original inception blocks are modified by substituting different spatial dimensions (5x5, 3x3, and 1x1) with only a single dimension (3x3) followed by a 1x1 convolution to reduce computational complexity. This decouples the spatial and feature map channels, making the network computationally efficient (Chollet, 2017). Xception first maps the cross-channel correlations using 1x1 convolution, then maps the spatial correlations of each output channel independently. Thus, Xception yields an increase in the classification performance compared to Inception_V3 (Cho and Choi, 2020). It is divided into 14 modules and has 36 convolutional layers. All modules have linear residual connections except for the first and last ones. The original Xception model consists of three flows named as entry, middle, and exit.

In the original Xception model, the middle flow is composed of three separable convolution layers, repeated eight times. That is, a total of 24 separable convolution layers are utilized. Considering the fact that, some mathematical calculations can be avoided by reducing the number of layers in the middle flow since the multi-label medical dataset is small (Kapadia and Paunwala, 2021). This layer, in the proposed modified Xception model, is employed once reducing 21 separable convolution layers. In addition, the global average pooling layer and the two separable convolution layers are removed in the exit flow while the entry flow has not been modified. The soft attention module, focusing primarily on relevant areas of skin image, is applied over the middle and exit flows to benefit from more distinctive representation of both modalities' data.

2.2. Hybrid-meta branch

The hybrid-meta branch is designed to integrate features from the dermoscopy and clinical modality branches to learn important spatial and channel context and exploit the relationship among the two modalities.

The features of these two modalities are initially concatenated into a single feature map. This feature map is then fed simultaneously both to a soft attention module and to a path of multiple convolutional units. Each convolutional unit in this path consists of ReLU, separable convolution and batch normalization layers. Then, a soft attention module is employed to focus on relevant areas of skin lesions forming a new feature map, and it is vertically concatenated with the feature map of the parallel soft attention module as well as that of the meta-data of patients to jointly explore the complementary information of three different modalities. Finally, this feature set is fed into a flatten layer followed by a dense layer to form the final feature vector.

2.3. Soft attention module

Only a small percentage of pixels in an image are relevant with skin lesion region while the rest are related with various irrelevant artifacts such as veins and hair. Therefore, in this paper, soft attention module is used to focus more on these relevant areas to classify skin lesions. For this, corresponding feature maps are multiplied by large weights to credit relevant areas of the image. The proposed soft attention module is shown Fig. 3.

A feature tensor obtained from either the entry or the exit flow of the modified Xception network defines the input of the soft attention module. This feature tensor, $t \in R^{hxwxd}$, fed into K counts of 3D convolution layers. The output of the 3D convolution set forms a new feature map, $t' \in R^{hxwxK}$, that is further normalized using a softmax layer to generate K counts of attention map, $t'' \in R^{hxwxK}$, as represented by Eq. 1 (Shaikh et al., 2020). Multiple relevant locations, in this feature map, receive high importance scores as a result of this normalization procedure.

$$t'' = \mathrm{softmax}\left(t'\right) = \frac{\exp\left(\mathrm{t_{ij}}'\right)}{\sum_{i=1}^{w}\sum_{j=1}^{h}\exp\left(\mathrm{t_{ij}}'\right)} \tag{1}$$

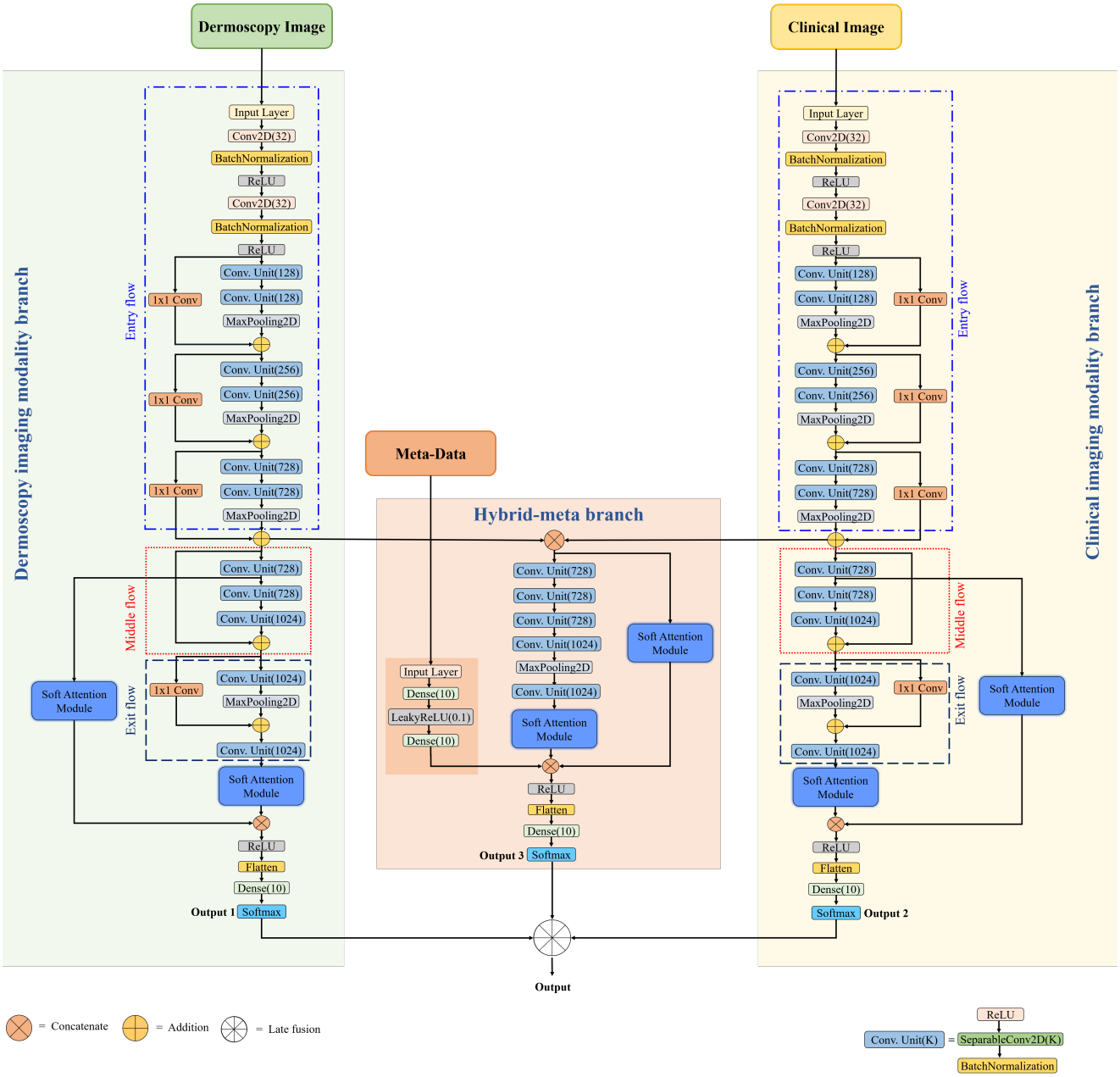


Fig. 2. The proposed multi-label skin lesion classification framework has three branches: dermoscopy imaging modality branch (green block), clinical imaging modality branch (yellow block), and a hybrid-meta branch (orange block). Modified Xception module based dermoscopy and clinical imaging modalities' features are first concatenated, then fed to the input of hybrid-meta branch, and finally concatenated with the meta-data.

Such obtained attention maps are aggregated to produce a unified attention map, represented by α, that acts as a weighting function (Eq. 2). It is then multiplied by t to attentively scale the salient feature values, and these features are also scaled with a learnable scalar parameter γ to form attentively scaled features (x_{sa}), as shown in Eq. 3, (Shaikh et al., 2020).

$$\alpha = \sum_{k=1}^{K} \text{softmax}\left(t'\right) \tag{2}$$

$$x_{sa} = \gamma t \alpha \tag{3}$$

Finally, x_{sa} concatenated with the original feature t in the form of a residual branch. The output of the soft attention module is given in Eq. 4 (Shaikh et al., 2020). This output helps to decide on the specific locations in an image for an effective feature map creation.

$$o = t + x_{sa} \tag{4}$$

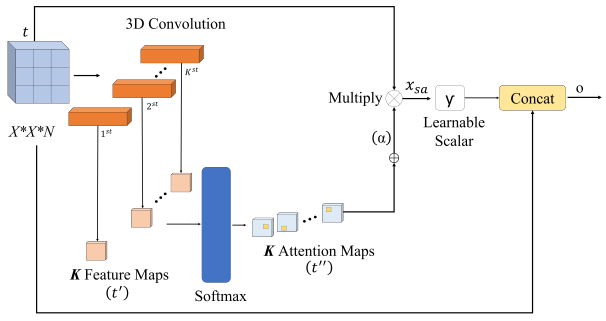


Fig. 3. The architecture of the soft attention module.

2.4. Fusion scheme

The proposed multi-modal network performs the final classification using Softmax classifier through a late fusion strategy. It is proposed to combine the decisions of multiple branches. Majority voting is one of the common late fusion approaches. This approach uses only the class labeling. It discards the probability information of the label, and thus, may lead to a performance loss. Therefore, average late fusion approach is proposed to utilize the class probability for multi-label skin lesion classification problem. That is, the class probabilities of the three branches are averaged, and the class with the highest average probability is assigned to every test case. The trained network produces prediction probabilities of the test cases, for each branch. For that, let m=1, 2 and 3 denote feature sets produced by dermoscopy, clinical and hybrid-meta branches, respectively. The corresponding probability set $P_{m,k}$ for $k = \{1, 2, 3 \ldots 8\}$ represents the prediction probability of mth branch for kth label. The averaged probability, P_k, is then obtained by averaging through all labels as shown in Eq. 5. The detection rates of each class are obtained by comparing the averaged prediction labels of the test cases with the actual classes.

$$P_k = \frac{1}{3} \sum_{m=1}^{3} P_{m,k} \tag{5}$$

3. Experimental Setup

This section provides information regarding (i) description of the dataset used during the experiments, (ii) evaluation criteria used to evaluate the proposed method, and (iii) details about the training process.

3.1. Dataset

The proposed method is evaluated on publicly available the Seven-Point Criteria Evaluation (SPC) dataset, which consists of dermoscopy image, clinical image and meta-data from 1,011 patients (Kawahara et al., 2018). The SPC dataset is split into official train, validation and test sets with 413, 203, and 395 cases, respectively. Distributation of samples among multiple labels(tasks) along with sub-labels for train, test and validation sets are shown in Table 1. Each case comprises a dermoscopy and a clinical image pairs, meta-data, the SPC labels and a diagnosis label, as seen in Fig. 1. Dermoscopy image sizes vary from 474 × 512 to 532 × 768 pixels and clinical image sizes vary from 480 × 512 to 532 × 768 pixels.

Table 1
The detailed statistics for 7-point criteria evaluation dataset.

Task	Category	Train	Validation	Test	Total
PN	ABS	160	84	156	400
	TYP	160	75	146	381
	ATP	93	44	93	230
BWV	ABS	339	157	320	816
	PRS	74	46	75	195
VS	ABS	347	163	313	823
	REG	43	22	52	117
	IR	23	18	30	71
PIG	ABS	253	112	223	588
	REG	44	26	48	118
	IR	116	65	124	305
STR	ABS	273	123	257	653
	REG	39	24	44	107
	IR	101	56	94	251
DAG	ABS	84	45	100	229
	REG	156	60	118	334
	IR	173	98	177	448
RS	ABS	317	152	289	758
	PRS	96	51	106	253
DIAG	BCC	19	7	16	42
	NEV	256	100	219	575
	MEL	90	61	101	252
	MISC	32	25	40	97
	SK	16	10	19	45

The SPC labels have 7 categories as Black Whitish Veil (BWV), Dots and Globules (DAG), Pigment (PIG), Pigment Network (PN), Regression Structures (RS), Streak (STR), and Vascular Structures (VS). Each label has different types, including Present (PRS), Absent (ASB), Typical (TYP), Atypical (ATY), Regular (REG), and Irregular (IRG). Also, there are five types of diagnostic labels in the SPC dataset. These are Basal Cell Carcinoma (BCC), Nevus (NEV), Melanoma (MEL), Miscellaneous (MISC) Seborrheic Keratosis (SK). It should be noted that samples in SPC dataset is unevenly distributed among all labels, and this is a major drawback in classification problems.

3.2. Evaluation Metrics

In order to compare the performance of the proposed method with other multi-modal learning studies using the same 7SPC dataset, namely (Alzahrani et al., 2019, Ge et al.,

Table 2
Comparison of classification results for single and multi-modality learning based on the proposed method without soft attention module in terms of accuracy (%).

Modality	BWV	DAG	PIG	PN	RS	STR	VS	DIAG	AVG
CIM	84.05 ± 3.6	74.43 ± 4.3	70.89 ± 4.5	57.47 ± 4.9	78.23 ± 4.1	68.61 ± 4.6	84.81 ± 3.5	59.75 ± 4.8	72.28 ± 4.4
DIM	85.82 ± 3.4	69.62 ± 4.5	74.18 ± 4.3	78.99 ± 4.0	**80.25** ± 3.9	65.32 ± 4.7	86.33 ± 3.4	69.11 ± 4.6	76.20 ± 4.2
HMM	**86.08** ± 3.4	60.51 ± 4.8	75.95 ± 4.2	77.97 ± 4.1	76.20 ± 4.2	**78.73** ± 4.0	**89.37** ± 3.0	42.53 ± 4.9	73.42 ± 4.4
DCIM	82.28 ± 3.8	**75.70** ± 4.2	75.19 ± 4.3	75.70 ± 4.2	78.23 ± 4.1	71.65 ± 4.4	86.33 ± 3.4	**75.19** ± 4.3	77.53 ± 4.1
DCIHMM	**86.08** ± 3.4	**75.70** ± 4.2	**76.96** ± 4.2	**80.51** ± 3.9	78.23 ± 4.1	76.20 ± 4.2	88.61 ± 3.1	**75.19** ± 4.3	**79.68** ± 4.0

Table 3 DÜZELT VS
Comparison of classification results for single and multi-modality learning based on the proposed method with soft attention module in terms of accuracy (%).

Modality	BWV	DAG	PİG	PN	RS	STR	VS	DİAG	AVG
CIM	84.30 ± 3.6	47.34 ± 4.9	73.92 ± 4.3	61.27 ± 4.8	80.76 ± 3.9	70.89 ± 4.5	81.27 ± 3.9	64.81 ± 4.7	70.57 ± 4.5
DIM	89.62 ± 3.0	74.94 ± 4.3	77.97 ± 4.1	72.15 ± 4.4	81.01 ± 3.9	76.96 ± 4.2	92.66 ± 2.6	72.41 ± 4.4	79.72 ± 4.0
HMM	85.32 ± 3.5	62.28 ± 4.8	70.63 ± 4.5	80.00 ± 3.9	77.47 ± 4.1	76.71 ± 4.2	87.85 ± 3.2	65.06 ± 4.7	75.66 ± 4.2
DCIM	86.58 ± 3.4	75.44 ± 4.2	77.47 ± 4.1	77.47 ± 4.1	**81.77** ± 3.8	78.99 ± 4.0	**96.20** ± 1.9	75.95 ± 4.2	81.23 ± 3.9
DCIHMM	**91.39** ± 2.8	**76.71** ± 4.2	**78.23** ± 4.1	**81.52** ± 3.8	81.52 ± 3.8	81.33 ± 3.8	95.44 ± 2.1	**78.23** ± 4.1	**83.04** ± 3.7

2017, He et al., 2023, Kawahara et al., 2018, Tang et al., 2022, Wang et al., 2021b, Yap et al., 2018), their adopted evaluation criteria are also employed in the current study. In this sense, sensitivity (SEN), specificity (SPE), precision (PRE) and the area under the curve (AUC) metrics are used to validate the effectiveness of the proposed method and make quantitative comparisons.The Receiver Operating Characteristic (ROC) curve plots sequence of SEN versus (1-SPEC) values. The area under the curve (AUC), on the other hand, is a global measure for discriminating whether a given condition in a test is present or not. To compare the performance of proposed model with other multi-modal learning studies, the average accuracy (AVG) is utilized as the main metric for the comparison. The definitions of accuracy, sensitivity, specificity and precision are as follows:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$Sen = \frac{TP}{TP + FN} \tag{7}$$

$$Spec = \frac{TN}{FP + TN} \tag{8}$$

$$Prec = \frac{TP}{TP + FP} \tag{9}$$

The confidence interval (CI) of the obtained accuracy rates can be calculated using the formula (Babukarthik et al., 2020, Minaee et al., 2020, Moung et al., 2021, Raschka, 2018, Sarkar et al., 2021).:

$$CI = Acc \pm z\sqrt{\frac{Acc(1 - Acc)}{n}} \tag{10}$$

where z is a constant value (1.96) corresponding to 95% confidence level, n is the size of the testing dataset.

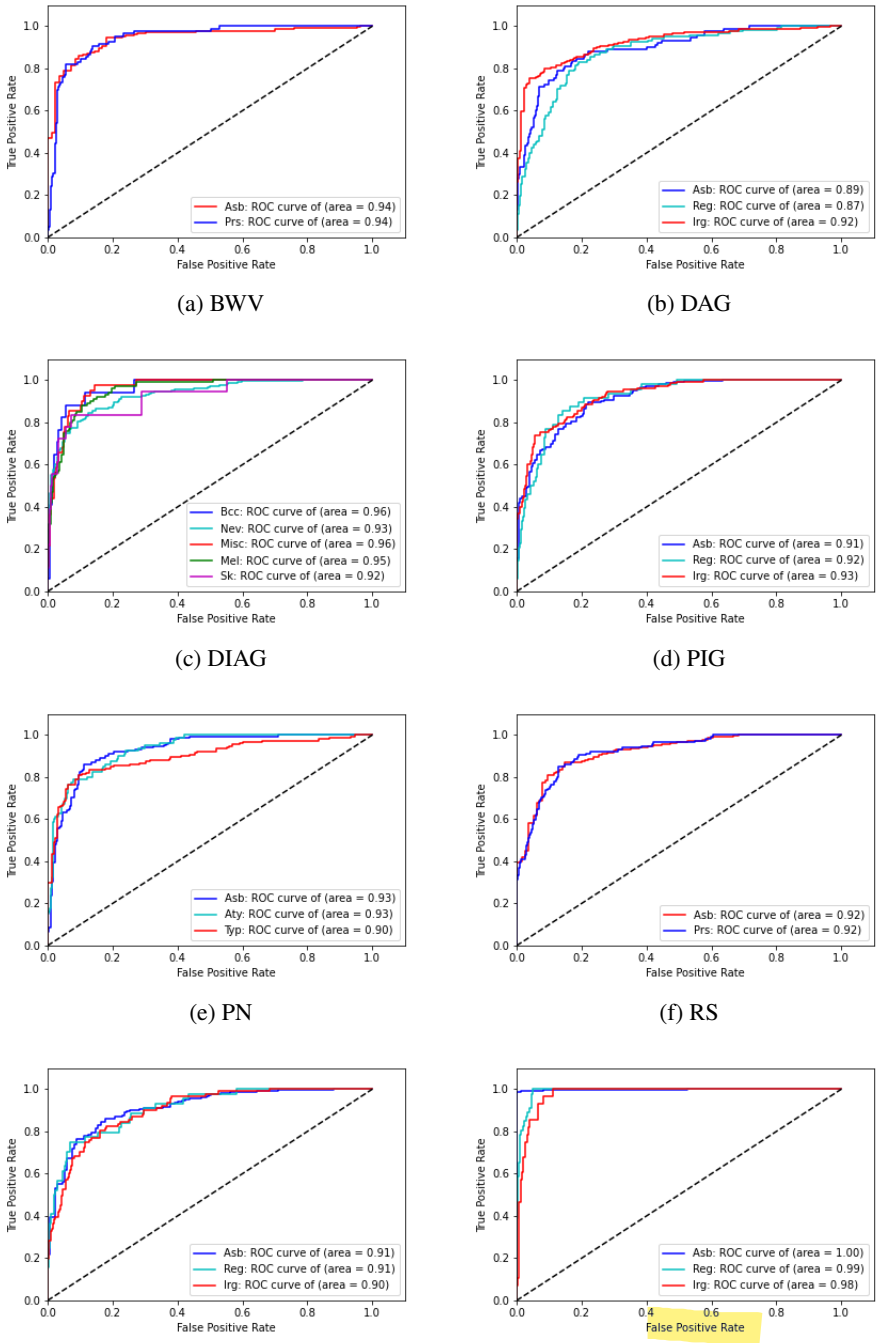
3.3. Implementation details

The following implementation and hyperparameter choices are used for all experiments. All images are resized to 299×299×3, the standard size for the XceptionNet. Experiments were implemented in Python using the Keras package with Tensorflow as backend and run on a 3x NVIDIA RTX 2080 Ti GPU with 12GB RAM hardware. The proposed network is trained end-to-end from scratch for 150 epochs with a batch size of 32. The proposed architecture contains about 22.721M trainable parameters to be updated during the implementation. The training stops after 150 epochs and the weights with the best accuracy on the validation dataset are saved for testing. During the training of proposed network, order of training examples is shuffled for every epoch. Adam optimizer with 0.0001 initial learning rate was used. Dropout in the last fully connected layer was used set to a rate of 0.5 to prevent overfitting.

4. Experimental results

In the proposed study, two sets of experiments were conducted. In the first, a modified Xception model was used for the multi-label skin lesion classification, while in the second, this model was evolved with the use of soft attention module to improve the classification performance. The corresponding results are presented in Table 2 and Table 3, respectively. Both tables compare the performance between single and multi-modality models in terms of the classification accuracy of each SPC labels, diagnosis (DIAG), and their average (AVG).

Table 2 shows the classification results that are trained by dermoscopy image modality (DIM) only and clinical image modality (CIM) only along with multi-modal experiments' results. These are late fusion of two image modalities (DCIM), hybrid multi-modality (HMM), that is patient's meta-data fused with other two imaging modalities, and finally the proposed late fusion of DIM, CIM and HMM, represented by DCIHMM. The obtained test results show an improvement in classification accuracy of multi-modality experiments compared to single-modality ones. In addition,



(a) BWV (b) DAG

(c) DIAG (d) PIG

(e) PN (f) RS

(g) STR (h) VS DÜZELT VS

Fig. 4. ROC curves for all 7-point checklist and diag label using soft attention module.

the proposed DCIHMM model improves the prediction performance for each SPC labels up to 5% compared to the DCIM model, which is widely used in the literature.

Obtained test results using the proposed method with soft attention module are presented in Table 3. It is clear from the comparison of Table 2 and Table 3 that the proposed method with soft attention module consistently improves the performance. In terms of the classification accuracy, each checklist label performance increases in the range of 1.25%– 7.71% while the diagnosis accuracy improved 4%. The overall average accuracy, on the other hand, improves 4.21%.

With the use of the proposed framework employing soft attention module, the receiver operating characteristic curve (ROC) for each SPC label along with diagnosis label are presented in Fig. 4. To better highlight the performance improvement, only the accuracy and AUC results are presented

in this section while the SEN, SPE and PRE results are shown and compared with the literature in next section.

5. Discussion

In this study, a new soft attention-based multi-modal learning network is proposed for multi-label skin lesion classification. The main findings of the proposed method are as follows. (i) The proposed network effectively learns complementary features from different modalities consisting of clinical images, dermoscopy images, and patients' metadata; (ii) soft attention module focuses on important lesion regions; (iii) the proposed method produces consistently high results for all labels considering the fact that no data balancing technique has been employed as opposed to most of the existing studies. The evaluation of the proposed network is presented in three subsections as follows.

5.1. Comparison between single and multi-modality learning

DÜZELT VS

The classification results are compared for different models based on single and multi-modality data to validate the effectiveness of multi-modal learning. The corresponding results are shown in Fig. 5. In single-modality learning, the average classification accuracy of 79.72% using dermoscopy image modality outperforms 70.57% accuracy of the model using clinical image modality, that is, about 13% improvement indicates the superiority of dermoscopy imaging modality compared to clinical imaging modality. On the other hand, all multi-modal learning models yield better performance than models using single modality as also illustrated by Fig. 5. For example, the model using multiple modalities of clinical and dermoscopy images achives higher accuracy with an improvement of 1.9% and 15.10% compared to dermoscopy and clinical modalities, respectively. These results indicate that the fusion of different modalities benefits in the classification performance.

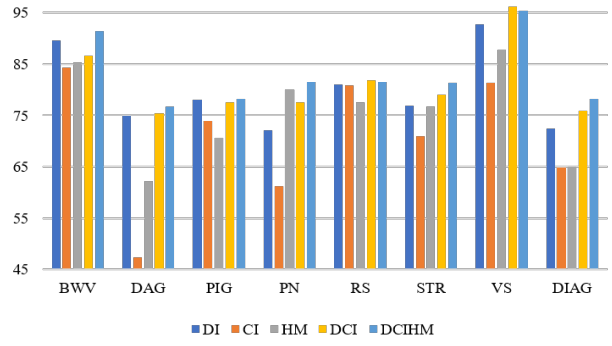


Fig. 5. Comparison of classification accuracy between single and multi-modal learning.

It has been also observed that adding the patients' metadata into the multiple imaging model in a hybrid manner can significantly improve the classification performance. That is, the proposed hybrid-based fusion approach improved the classification accuracy by an average of 7.36% for all label types, and multi-modal learning provides significant as well as complementary information from different modalities for multi-label skin lesion classification. Hence, the proposed method can achieve a promising performance for skin lesion classification problem.

5.2. Effectiveness of soft attention module

DÜZELT VS

An ablation study was performed to determine the efficacy of the proposed soft attention module. Fig. 6 shows the corresponding results for all labels with and without the use of soft attention modules. It has been observed that the use of proposed soft attention modules improves the average accuracy from 79.68% to 83.04% with an improvement of 4.2%. That is, use of soft attention module enables better feature representation for all modalities as it focuses on relevant areas of images and guarantees better representations in multi-label skin lesion classification problem.

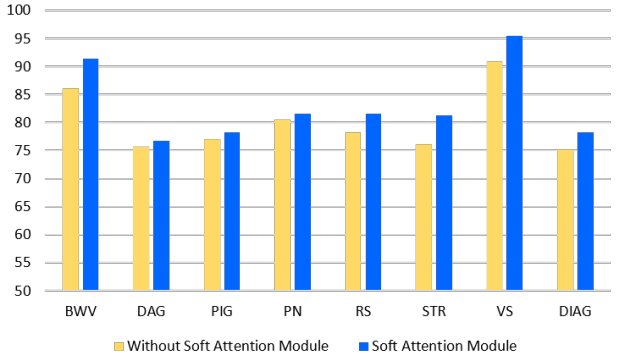


Fig. 6. Comparison of classification accuracy with/without soft attention module.

Finally, in Fig. 7, different dataset images along with their corresponding heat map figures, obtained from last soft attention layers, are presented to illustrate the performance of soft attention-based hybrid deep learning model on focusing lesion areas. It improves the training performance and results in improved accuracy of the proposed model for skin lesion classification problem.

5.3. Comparison with the state-of-the-art methods

The performance of the proposed method is compared with the current state-of-the-art methods in terms of accuracy in Table 4. These methods all used the same dataset, same classification task, and same train-validation-test partitions for all seven-checklist criteria detection problem. Based on the experimental results of Table 4, (Alzahrani et al., 2019) and (Ngiam et al., 2011) yielded the lowest average accuracy among all studies. This is because their models learn from a single shared layer, which are not capable of capturing important spatial information from the dermoscopy image. TripleNet (Ge et al., 2017) achieves a 1.4% higher average accuracy than EmbeddingNet (Yap et al., 2018) due to the efficiency of the extra subnet and the combined optimization of single and cross-modality.

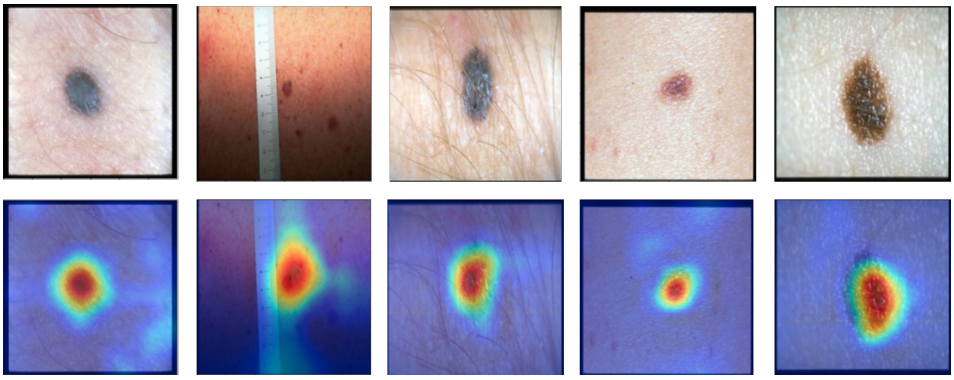


Fig. 7. Visualization results for evaluating the effectiveness of the proposed method based on the soft attention module.

Table 4
Comparison of classification results to the state-of-the-art methods in terms of accuracy (%).

Methods	BWV	DAG	PIG	PN	RS	STR	VS	DIAG	AVG
Alzahrani et al. (Alzahrani et al., 2019)	80.5	50.4	62.8	62.8	71.9	67.6	73.2	64.3	66.7
MmA (Ngiam et al., 2011)	83	59.2	61.3	65.6	73.9	69.4	75.7	70.6	69.8
EmbeddingNet (Yap et al., 2018)	84.3	57.5	64.3	65.1	78.0	73.4	82.5	68.6	71.7
TripleNet (Ge et al., 2017)	87.9	61.3	67.3	63.3	76.0	74.4	83.0	68.6	72.7
Kawahara et al. (Kawahara et al., 2018)	87.1	60.0	66.1	70.9	77.2	74.2	79.7	74.2	73.7
HcCNN (Bi et al., 2020)	87.1	69.9	68.6	70.6	80.8	71.6	84.8	69.9	74.9
CAFNet (He et al., 2023)	87.8	61.5	73.4	70.1	81.8	77.0	84.3	78.2	76.8
FusionM4Net-SS (Tang et al., 2022)	88.1	66.1	70.1	71.1	81.5	78.0	81.8	**78.5**	77.0
CC (Wang et al., 2021b)	88.1	75.7	71.6	81.0	81.5	**81.3**	93.9	77.5	81.3
Proposed	**91.4**	**76.7**	**78.2**	**81.5**	**81.5**	**81.3**	**95.4**	78.2	**83.04**

Although the Inception-combined method (Kawahara et al., 2018) obtains higher accuracy than HcCNN (Bi et al., 2020) for PN, STR and DIAG labels, the latter has better performance for other labels including the average accuracy with 1.6% higher performance. This results from the fact that hybrid network structure of HcCNN fuses the feature information of CNN backbone for different stages. Considering the up-to-date literature, the CC based method (Wang et al., 2021b), a combined framework of multi-labeled deep feature extractor with clinically constrained classifier chain, acquires the highest average accuracy rate while FusionM4Net-SS yields the highest DIAG accuracy with 1.6% lower average accuracy than that of the (Wang et al., 2021b). The proposed study, on the other hand, achieves the highest accuracy scores for all labels including average accuracy except for DIAG label. Compared with the FusionM4Net, the proposed method has a significant increase of 2.14% in average accuracy, utilizing hybrid fusion of three-modality data (clinical image, dermoscopy image, and patient's meta-data). The performance comparison results,Table 4, shows the superiority of the proposed multi-branch, multi-modal learning algorithm that effectively extracts and fuses information from two-modality image data first, then exploits the information of the patients' meta-data on the hybrid domain.

Table 5 further compares the performance of the Inception-combined, FusionM4Net-SS algorithms and the proposed study. In the table, the highest value of SEN, SPE, PRE and AUC metrics are highlighted in color red, blue, green and black, respectively. Based on the Table 5 results, the proposed method achieves the highest AUC value for all label types. Its SEN, SPE, and PRE metrics performance, on the other hand, is not that certain with some missing entries, about 30% of these three labels, while the proposed method dominantly performs better. It is considered that the imbalanced distribution of the data among all label types, as shown in Table 1, is responsible for this performance loss. That is, less unbalanced label types, such as DAG and BWV, provided better SEN, SPE, and PRE performance compared to highly unbalanced label types. For example, SEN score of DIAG is 17.6%. In other words, the remaining 82.4% of corresponding 16 test-samples out of 395 samples are misclassified resulting in a degradation of DIAG label accuracy score, which is slightly less than the highest. Still, the proposed attention-based model produces the highest scores for all other labels including AVG.

6. Conclusion

In this study, a soft attention-based multi-modal learning algorithm is proposed for multi-label skin lesion classification to leverage multiple modalities of medical data. The proposed method introduced a deep hybrid-connected convolutional neural network with a soft attention module to integrate the visual characteristics of two-modality images and patients' meta-data. The proposed method can learn

Table 5 DÜZELT VS

The detailed seven-point checklist label and diagnosis label classification results in terms of sensitivity (SEN), specificity (SPE), precision (PRE) and AUC (%).

		BWV		DAG			PIG			PN			RS		STR			VS			DIAG				
Model	Metric	ABS	PRS	ABS	REG	IRRE	ABS	REG	IRRE	ASB	ATY	TYP	ABS	PRS	ABS	REG	IRRE	ABS	REG	IRRE	BCC	NEV	MISC	MEL	SK
Inception	SEN	89.4	77.3	47	67.8	62.1	77.6	29.2	59.7	77.6	48.4	78.1	81.3	66	86	54.5	51.1	92.3	42.3	13.3	62.5	88.6	47.5	61.4	42.1
	SPE	77.3	89.4	87.8	72.6	78.9	65.1	94.2	80.1	85.8	90.7	78.7	66	81.3	67.4	85.7	85.7	45.1	92.4	97.5	97.9	71.6	97.5	88.8	99.5
	PRE	94.4	63	56.6	51.3	70.5	74.2	41.2	57.8	78.1	61.6	68.3	86.7	56.5	83.1	52.7	52.7	86.5	45.8	30.8	55.6	79.5	67.9	65.3	80
	AUC	89.2	89.2	74.1	76.5	79	79	74.9	79	89.9	79.9	84.2	82.9	82.9	86.1	78.9	78.9	86.2	85.5	76.1	92.9	89.7	88.3	86.3	91
FusionM4Net-SS	SEN	94.4	64	40 .0	76.3	76.3	89.7	18.8	54.8	76.3	49.5	79.5	95.2	44.3	91.4	47.7	55.3	98.1	28.8	0.3	43.8	95	52.5	71.3	10.5
	SPE	64	94.4	92.5	78	76.6	50	98.8	89.7	83.3	90.7	81.5	44.3	95.2	66.7	97.2	89.7	29.3	96.2	99.7	97.9	73.3	98.9	91.5	99.7
	PRE	91.8	72.7	64.5	58.5	72.6	69.9	69.2	70.8	74.8	62.2	71.6	82.3	77	83.6	67.7	62.7	84.1	53.6	50	46.7	81.6	84	74.2	66.7
	AUC	92.5	92.5	78.2	82.5	84.5	83.1	82.5	84.9	88.1	85.7	87.3	83	83	87.2	89.6	84.6	89.5	87.9	81.3	95.4	94.6	95	92.6	91.5
Proposed	SEN	97.1	69.9	63.3	81.4	80.2	90.3	27.1	76.9	89	66.3	81.5	98.9	38.6	90.8	54.5	67	99.4	82.6	71.4	17.6	82.6	90.2	75	72.2
	SPE	69.9	97.1	94.1	81.9	88.5	71.9	97.7	89.4	85.7	95.9	88.9	38.6	98.9	72.6	96.6	91.8	93.2	98.3	98.1	99.2	89.2	87.9	94.9	98.4
	PRE	92.4	86.6	76	65.8	86.2	79.7	61.9	78.1	81.6	80.3	82	79.9	93.6	86.4	66.7	70.9	98.5	86.4	74.1	50	90.5	46.3	83.3	68.4
	AUC	**94.4**	**94.4**	**89.5**	**87.5**	**92.1**	**91.3**	**91.9**	**92.6**	**92.9**	**93.2**	**89.7**	**91.9**	**91.9**	**91.1**	**91**	**90**	**99.8**	**99**	**97.7**	**96.2**	**93.1**	**96.1**	**95.1**	**92.1**

complementary information from different modalities with dermoscopy branch, clinical branch and hybrid-meta branch. The proposed dermoscopy and clinical imaging modality branches employ modified Xception as the backbone. The hybrid-meta branch, which fuses feature maps derived from the two imaging modalities, uses multiple convolution layers to learn complementary information. Specifically, the hybrid-meta branch combines features of three different modalities in a hybrid manner to obtain complementary information. All proposed branches uses soft attention modules to make the CNN backbones pay more attention to the lesion for better feature extraction. The comprehensive experiments and comparisons on a publicly available dataset, 7-point criteria evaluation database, demonstrated that the proposed method has higher accuracy than other current state-of-the-art methods, with an average accuracy improvement of 2.41%. Obtained results show that the proposed method is able to achieve consistent and the best classification results in datasets with imbalanced label distributions. Although proposed network may suffer from some hands-on limitations just like similar studies such as the lack of chronologically sorted clinical and/or dermoscopy images, obtained network parameters not being efficiently transferable to dataset of different ethnic origins and possibility of having noisy images, this network defines a new basis for the existing skin lesion classification problem. Use of a graph convolutional network (GCN) with an attention module is planned as the future work to capture the valuable label dependencies and further improve the performance of multi-label skin lesion classification problem.

Our code will be available at https://github.com/AsHuTeam/Multi-Label-Skin-Lesion-Classification.git.

Acknowledgement

This work is funded by Ataturk University Scientific Research Projects Grant FBA-2022-11555.

References

Abdelhalim, I.S.A., Mohamed, M.F., Mahdy, Y.B., 2021. Data augmentation for skin lesion using self-attention based progressive generative adversarial network. Expert Systems with Applications 165, 113922.

Adegun, A., Viriri, S., 2021. Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. Artificial Intelligence Review 54, 811–841.

Alzahrani, S., Al-Nuaimy, W., Al-Bander, B., 2019. Seven-point checklist with convolutional neural networks for melanoma diagnosis, in: 2019 8th European Workshop on Visual Information Processing (EUVIP), IEEE. pp. 211–216.

Antonelli, M., Ducange, P., Marcelloni, F., 2014. An experimental study on evolutionary fuzzy classifiers designed for managing imbalanced datasets. Neurocomputing 146, 125–136.

Argenziano, G., Catricalà, C., Ardigo, M., Buccini, P., De Simone, P., Eibenschutz, L., Ferrari, A., Mariani, G., Silipo, V., Sperduti, I., et al., 2011. Seven-point checklist of dermoscopy revisited. British Journal of Dermatology 164, 785–790.

Arora, G., Dubey, A.K., Jaffery, Z.A., Rocha, A., 2020. Bag of feature and support vector machine based early diagnosis of skin cancer. Neural Computing and Applications , 1–8.

Babukarthik, R., Adiga, V.A.K., Sambasivam, G., Chandramohan, D., Amudhavel, J., 2020. Prediction of covid-19 using genetic deep learning convolutional neural network (gdcnn). Ieee Access 8, 177647–177666.

Bai, J., Ding, B., Xiao, Z., Jiao, L., Chen, H., Regan, A.C., 2021. Hyperspectral image classification based on deep attention graph convolutional network. IEEE Transactions on Geoscience and Remote Sensing 60, 1–16.

Balch, C.M., Gershenwald, J.E., Soong, S.j., Thompson, J.F., Atkins, M.B., Byrd, D.R., Buzaid, A.C., Cochran, A.J., Coit, D.G., Ding, S., et al., 2009. Final version of 2009 ajcc melanoma staging and classification. Journal of clinical oncology 27, 6199.

Benyahia, S., Meftah, B., Lézoray, O., 2022. Multi-features extraction based on deep learning for skin lesion classification. Tissue and Cell 74, 101701.

Bi, L., Feng, D.D., Fulham, M., Kim, J., 2020. Multi-label classification of multi-modality skin lesion via hyper-connected convolutional neural network. Pattern Recognition 107, 107502.

Brinker, T.J., Hekler, A., Enk, A.H., Klode, J., Hauschild, A., Berking, C., Schilling, B., Haferkamp, S., Schadendorf, D., Holland-Letz, T., et al., 2019. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. European Journal of Cancer 113, 47–54.

Cai, G., Zhu, Y., Wu, Y., Jiang, X., Ye, J., Yang, D., 2022. A multimodal transformer to fuse images and metadata for skin disease classification. The Visual Computer , 1–13.

Calderón, C., Sanchez, K., Castillo, S., Arguello, H., 2021. Bilsk: A bilinear convolutional neural network approach for skin lesion classification. Computer Methods and Programs in Biomedicine Update 1, 100036.

Carli, P., de Giorgi, V., Chiarugi, A., Nardini, P., Weinstock, M.A., Crocetti, E., Stante, M., Giannotti, B., 2004. Addition of dermoscopy to conventional naked-eye examination in melanoma screening: a randomized study. Journal of the American Academy of Dermatology 50, 683–689.

de Carvalho, T.M., Noels, E., Wakkee, M., Udrea, A., Nijsten, T., 2019. Development of smartphone apps for skin cancer risk assessment: progress and promise. JMIR Dermatology 2, e13376.

Chen, Q., Li, M., Chen, C., Zhou, P., Lv, X., Chen, C., 2022. Mdfnet: application of multimodal fusion method based on skin image and clinical data to skin cancer classification. Journal of Cancer Research and Clinical Oncology , 1–13.

Cheng, J.Z., Ni, D., Chou, Y.H., Qin, J., Tiu, C.M., Chang, Y.C., Huang, C.S., Shen, D., Chen, C.M., 2016. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. Scientific reports 6, 1–13.

Cho, W.K., Choi, S.H., 2020. Comparison of convolutional neural network models for determination of vocal fold normality in laryngoscopic images. Journal of Voice .

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251–1258.

Dhivyaa, C., Sangeetha, K., Balamurugan, M., Amaran, S., Vetriselvi, T., Johnpaul, P., 2020. Skin lesion classification using decision trees and random forest algorithms. Journal of Ambient Intelligence and Humanized Computing , 1–13.

Ding, S., Wu, Z., Zheng, Y., Liu, Z., Yang, X., Yang, X., Yuan, G., Xie, J., 2021. Deep attention branch networks for skin lesion classification. Computer methods and programs in biomedicine 212, 106447.

Fisher, R.B., Rees, J., Bertrand, A., 2019. Classification of ten skin lesion classes: Hierarchical knn versus deep net, in: Annual Conference on Medical Image Understanding and Analysis, Springer. pp. 86–98.

Fu, X., Bi, L., Kumar, A., Fulham, M., Kim, J., 2022. Graph-based intercategory and intermodality network for multilabel classification and melanoma diagnosis of skin lesions in dermoscopy and clinical images. IEEE Transactions on Medical Imaging .

Ge, Z., Demyanov, S., Chakravorty, R., Bowling, A., Garnavi, R., 2017. Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images, in: International conference on medical image computing and computer-assisted intervention, Springer. pp. 250–258.

Gessert, N., Sentker, T., Madesta, F., Schmitz, R., Kniep, H., Baltruschat, I., Werner, R., Schlaefer, A., 2019. Skin lesion classification using cnns with patch-based attention and diagnosis-guided loss weighting. IEEE Transactions on Biomedical Engineering 67, 495–503.

Goyal, M., Knackstedt, T., Yan, S., Hassanpour, S., 2020. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. Computers in Biology and Medicine 127, 104065.

Haenssle, H.A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A.B.H., Thomas, L., Enk, A., et al., 2018. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of oncology 29, 1836–1842.

Hasan, M.K., Elahi, M.T.E., Alam, M.A., Jawad, M.T., Martí, R., 2022. Dermoexpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation. Informatics in Medicine Unlocked 28, 100819.

Hasan, M.K., Roy, S., Mondal, C., Alam, M.A., Elahi, M.T.E., Dutta, A., Raju, S.T.U., Jawad, M.T., Ahmad, M., 2021. Dermo-doctor: A framework for concurrent skin lesion detection and recognition using a deep convolutional neural network with end-to-end dual encoders. Biomedical Signal Processing and Control 68, 102661.

Hassanzadeh, T., Essam, D., Sarker, R., 2021. Evolutionary deep attention convolutional neural networks for 2d and 3d medical image segmentation. Journal of Digital Imaging 34, 1387–1404.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

He, X., Wang, Y., Zhao, S., Chen, X., 2023. Co-attention fusion network for multimodal skin cancer diagnosis. Pattern Recognition 133, 108990.

Kapadia, M.R., Paunwala, C.N., 2021. Content based medical image retrieval system for accurate disease diagnoses using modified multi feature fused xception model. Indian Journal of Computer Science and Engineering (IJCSE) 12, 89–100.

Kawahara, J., BenTaieb, A., Hamarneh, G., 2016. Deep features to classify skin lesions, in: 2016 IEEE 13th international symposium on biomedical imaging (ISBI), IEEE. pp. 1397–1400.

Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G., 2018. Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE journal of biomedical and health informatics 23, 538–546.

Khan, M.A., Akram, T., Zhang, Y.D., Sharif, M., 2021a. Attributes based skin lesion detection and recognition: A mask rcnn and transfer learning-based deep learning framework. Pattern Recognition Letters 143, 58–66.

Khan, M.A., Muhammad, K., Sharif, M., Akram, T., de Albuquerque, V.H.C., 2021b. Multi-class skin lesion detection and classification via teledermatology. IEEE Journal of Biomedical and Health Informatics 25, 4267–4275.

Khan, M.A., Zhang, Y.D., Sharif, M., Akram, T., 2021c. Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification. Computers & Electrical Engineering 90, 106956.

Li, W., Zhuang, J., Wang, R., Zhang, J., Zheng, W.S., 2020. Fusing metadata and dermoscopy images for skin disease diagnosis, in: 2020 IEEE 17th international symposium on biomedical imaging (ISBI), IEEE. pp. 1996–2000.

Liang, R., Wu, Q., Yang, X., 2019. Multi-pooling attention learning for melanoma recognition, in: 2019 Digital Image Computing: Techniques and Applications (DICTA), IEEE. pp. 1–6.

Liu, Y., Jain, A., Eng, C., Way, D.H., Lee, K., Bui, P., Kanada, K., de Oliveira Marinho, G., Gallegos, J., Gabriele, S., et al., 2020. A deep learning system for differential diagnosis of skin diseases. Nature medicine 26, 900–908.

Menegola, A., Fornaciali, M., Pires, R., Bittencourt, F.V., Avila, S., Valle, E., 2017. Knowledge transfer for melanoma screening with deep learning, in: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), IEEE. pp. 297–300.

Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., Soufi, G.J., 2020. Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. Medical image analysis 65, 101794.

Moung, E.G., Hou, C.J., Sufian, M.M., Hijazi, M.H.A., Dargham, J.A., Omatu, S., 2021. Fusion of moment invariant method and deep learning algorithm for covid-19 classification. Big Data and Cognitive Computing 5, 74.

Mporas, I., Perikos, I., Paraskevas, M., 2020. Color models for skin lesion classification from dermatoscopic images, in: Advances in Integrations of Intelligent Methods. Springer, pp. 85–98.

Nachbar, F., Stolz, W., Merkle, T., Cognetta, A.B., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O., Plewig, G., 1994. The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. Journal of the American Academy of Dermatology 30, 551–559.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y., 2011. Multimodal deep learning, in: ICML, pp. 689–696. URL: https://icml.cc/2011/papers/399_icmlpaper.pdf.

Panthakkan, A., Anzar, S., Jamal, S., Mansoor, W., 2022. Concatenated xception-resnet50—a novel hybrid approach for accurate skin cancer prediction. Computers in Biology and Medicine 150, 106170.

Pathan, S., Prabhu, K.G., Siddalingaswamy, P., 2018. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review. Biomedical Signal Processing and Control 39, 237–262.

Puig, S., Argenziano, G., Zalaudek, I., Ferrara, G., Palou, J., Massi, D., HOFMANN-WELLENHOF, R., Soyer, H.P., Malvehy, J., 2007. Melanomas that failed dermoscopic detection: a combined clinicodermoscopic approach for not missing melanoma. Dermatologic surgery 33, 1262–1273.

Raschka, S., 2018. Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:1811.12808 .

Roh, S.B., Pedrycz, W., Ahn, T.C., 2014. A design of granular fuzzy classifier. Expert Systems with Applications 41, 6786–6795.

Sarkar, A., Vandenhirtz, J., Nagy, J., Bacsa, D., Riley, M., 2021. Identification of images of covid-19 from chest x-rays using deep learning: comparing cognex visionpro deep learning 1.0™ software with open source convolutional neural networks. SN Computer Science 2, 1–16.

Shaikh, M.A., Duan, T., Chauhan, M., Srihari, S.N., 2020. Attention based writer independent verification, in: 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE. pp. 373–379.

Sharafudeen, M., et al., 2022. Detecting skin lesions fusing handcrafted features in image network ensembles. Multimedia Tools and Applications , 1–21.

Sharma, A.K., Tiwari, S., Aggarwal, G., Goenka, N., Kumar, A., Chakrabarti, P., Chakrabarti, T., Gono, R., Leonowicz, Z., Jasiński, M., 2022. Dermatologist-level classification of skin cancer using cascaded ensembling of convolutional neural network and handcrafted features based deep neural network. IEEE Access 10, 17920–17932.

Siegel, R.L., Miller, K.D., Jemal, A., 2019. Cancer statistics, 2019. CA: a cancer journal for clinicians 69, 7–34.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Sitaula, C., Hossain, M.B., 2021. Attention-based vgg-16 model for covid-19 chest x-ray image classification. Applied Intelligence 51, 2850–2863.

Soyer, H.P., Argenziano, G., Zalaudek, I., Corona, R., Sera, F., Talamini, R., Barbato, F., Baroni, A., Cicale, L., Di Stefani, A., et al., 2004. Three-point checklist of dermoscopy. Dermatology 208, 27–31.

Stojanovic, V., Nedic, N., 2016. Robust kalman filtering for nonlinear multivariable stochastic systems in the presence of non-gaussian noise. International Journal of Robust and Nonlinear Control 26, 445–460.

Tang, P., Liang, Q., Yan, X., Xiang, S., Sun, W., Zhang, D., Coppola, G., 2019. Efficient skin lesion segmentation using separable-unet with stochastic weight averaging. Computer methods and programs in biomedicine 178, 289–301.

Tang, P., Liang, Q., Yan, X., Xiang, S., Zhang, D., 2020. Gp-cnn-dtel: Global-part cnn model with data-transformed ensemble learning for skin lesion classification. IEEE journal of biomedical and health informatics 24, 2870–2882.

Tang, P., Yan, X., Nan, Y., Xiang, S., Krammer, S., Lasser, T., 2022. Fusionm4net: A multi-stage multi-modal learning algorithm for multi-label skin lesion classification. Medical Image Analysis 76, 102307.

Tao, H., Li, J., Chen, Y., Stojanovic, V., Yang, H., 2020. Robust point-to-point iterative learning control with trial-varying initial conditions. IET Control Theory & Applications 14, 3344–3350.

Tschandl, P., Rosendahl, C., Akay, B.N., Argenziano, G., Blum, A., Braun, R.P., Cabo, H., Gourhant, J.Y., Kreusch, J., Lallas, A., et al., 2019. Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. JAMA dermatology 155, 58–65.

Versaci, M., Angiulli, G., Crucitti, P., De Carlo, D., Laganà, F., Pellicanò, D., Palumbo, A., 2022. A fuzzy similarity-based approach to classify numerically simulated and experimentally detected carbon fiber-reinforced polymer plate defects. Sensors 22, 4232.

Vestergaard, M., Macaskill, P., Holt, P., Menzies, S., 2008. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. British Journal of Dermatology 159, 669–676.

Wang, S., Yin, Y., Wang, D., Wang, Y., Jin, Y., 2021a. Interpretability-based multimodal convolutional neural networks for skin lesion diagnosis. IEEE Transactions on Cybernetics .

Wang, Y., Cai, J., Louie, D.C., Wang, Z.J., Lee, T.K., 2021b. Incorporating clinical knowledge with constrained classifier chain into a multimodal deep network for melanoma detection. Computers in Biology and Medicine 137, 104812.

Wang, Y., Feng, Y., Zhang, L., Zhou, J.T., Liu, Y., Goh, R.S.M., Zhen, L., 2022. Adversarial multimodal fusion with attention mechanism for skin lesion classification using clinical and dermoscopic images. Medical Image Analysis 81, 102535.

Wu, J., Hu, W., Wen, Y., Tu, W., Liu, X., 2020. Skin lesion classification using densely connected convolutional networks with attention residual learning. Sensors 20, 7080.

Yan, Z., Liu, W., Wen, S., Yang, Y., 2019. Multi-label image classification by feature attention network. Ieee Access 7, 98005–98013.

Yap, J., Yolland, W., Tschandl, P., 2018. Multimodal skin lesion classification using deep learning. Experimental dermatology 27, 1261–1267.

Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A., 2016. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE transactions on medical imaging 36, 994–1004.

Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., Lei, B., Wang, T., 2018. Melanoma recognition in dermoscopy images via aggregated deep convolutional features. IEEE Transactions on Biomedical Engineering 66, 1006–1016.

Zhang, J., Xie, Y., Xia, Y., Shen, C., 2019. Attention residual learning for skin lesion classification. IEEE transactions on medical imaging 38, 2092–2103.

Zhang, X., Wang, H., Stojanovic, V., Cheng, P., He, S., Luan, X., Liu, F., 2021. Asynchronous fault detection for interval type-2 fuzzy nonhomogeneous higher-level markov jump systems with uncertain transition probabilities. IEEE Transactions on Fuzzy Systems .

Zhou, C., Tao, H., Chen, Y., Stojanovic, V., Paszke, W., 2022. Robust point-to-point iterative learning control for constrained systems: A minimum energy approach. International Journal of Robust and Nonlinear Control 32, 10139–10161.