

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Early Detection of Multiclass Skin Lesions Using Transfer Learning-based IncepX-Ensemble Model

SUBHAJIT CHATTERJEE¹, JOON-MIN GIL², AND YUNG-CHEOL BYUN³

¹Jeju National University, Department of Computer Engineering, South Korea, Jeju-63243 (e-mail: subhajitchatterjee@stu.jejunu.ac.kr)

²Department of Computer Engineering, Jeju National University, Jeju 63243, South Korea (e-mail: jmgil@jejunu.ac.kr)

³Department of Computer Engineering, Major of Electronic Engineering, Jeju National University, Institute of Information Science & Technology, Jeju 63243, South Korea (e-mail: ycb@jejunu.ac.kr)

Corresponding author: Yung-Cheol Byun (e-mail: ycb@jejunu.ac.kr).

This result was supported by the "Regional Innovation Strategy(RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE).

ABSTRACT Skin lesion diagnosis in medical image analysis is still a difficult task. A frequent kind of cancer known as skin cancer affects people worldwide and can be fatal. As a result, early and accurate diagnosis is crucial for finding skin cancer patients. One of the most recent technologies for detecting skin cancer is dermatoscopy. For proper treatment of skin cancer, an early diagnosis is required. The early stages of skin lesions are identical, making manual diagnosis difficult. Therefore, creating computer-aided methods for classifying skin lesions will assist dermatologists in detecting skin lesions earlier and treating them more successfully. Current research indicates a significant potential for the classification of skin lesions using deep learning networks. However, issues including unbalanced datasets, poor contrast lesions, and the extraction of pointless or duplicate features still need to be resolved. This study aims to propose a transfer learning-based ensemble model for more accurate classification results. InceptionV3 and Xception are employed to build an ensemble model for classifying skin lesion images known as IncepX-Ensemble. A traditional data augmentation method was employed for the HAM10000 dataset to address the imbalanced data. This technique mitigates the class imbalance by incorporating data augmentation, enhancing model accuracy. At the experiment's outset, we utilized the original imbalanced data, and subsequently, balanced data were employed for the proposed model. Training and testing accuracy was achieved at 86% on the imbalanced dataset. A balanced data augmentation dataset yielded 98% training and 98% test accuracy rates. We evaluated the output of the proposed model against outputs from various transfer learning models using both the original and balanced datasets.

INDEX TERMS Deep learning, skin lesion classification, convolutional neural network, transfer learning, ensemble learning, image classification

I. INTRODUCTION

THE human skin, consisting of multiple layers, stands as the body's most crucial and sun-exposed organ. Enhancing the skin's role as a protective barrier involves employing fluids to prevent the degradation of lipids in the epidermis. Various factors, including fungal growth on the skin, imperceptible bacteria, bacterial alterations affecting skin texture, or changes in pigmentation, can give rise to skin symptoms [1].

The abnormal growth of skin cells is another name for skin cancer, and skin lesion cases have recently sharply

increased worldwide [2]. Skin cancer is caused by exposure to U.V. light from the sun, not by a depletion in the ozone layer, which protects from those rays. Basal cell carcinoma, melanoma, squamous cell carcinoma, and actinic keratosis (solar keratosis) are some of the different types of skin cancer. Out of all the other types of skin cancer, melanoma is the most dangerous. Melanoma starts in the brown or black melanocyte cells [3], which resemble a mole. Melanoma cells penetrate deeper skin layers from the skin's surface. The cells (melanocytes) that create melanin, the pigment that gives your skin its color, and which are responsible for

melanoma, the most serious form of skin cancer, develops. Melanoma has a 7% incidence rate, yet it accounts for 75% of fatal cancer cases [4]. The head, neck, face, and hands are just a few body places where skin cancer can develop due to increased exposure to the sun's harmful rays. Skin lesions must be identified and treated as soon as possible for the patient to survive. Given the resemblance between skin cancer, it can be challenging for dermatologists with experience to determine a skin cancer class. However, studies suggest that even skilled dermatologists can detect skin cancer with unaided eyes and visually examine it with an accuracy of about 75%-80% [5], [6]. Dermatologists often utilize the ABCD rule to diagnose skin cancer. Asymmetry, border, color, and diameter are the components of the ABCD rule [7]. Physicians identified the following features in skin lesions during the diagnosis. Dermatologists' diagnostic procedures are quite individualized. Numerous elements, including low picture resolutions, less expertise, and visual impairment, can lead to discrepancies in dermatologists' diagnoses and outcomes. These issues can have less effect if computer-based diagnosis systems are developed to enable such diagnostic applications [8]. Researchers may encounter misleading the multiclass classification of skin cancer when the dataset is unbalanced. As a result, it's important to employ a data-balanced technique that will balance the classes, which is vital for the final classification. Due to the similarity of skin lesions and the unbalanced nature of the data, the multiclass classification problem is more difficult than the binary class classification problem. Because of the skin cancer differences in skin surfaces, identifying the exact place is difficult. As a result, even for seasoned dermatologists, manually identifying skin lesions via dermatoscopy takes time to determine and penetrate more error potential. Because of this, researchers emphasize various computer-aided diagnostic (CAD) methods based on machine learning. Dermatologists use computerized dermatoscopy examination to transmit images of skin lesion samples taken from the human body to the computer and assess the images to determine whether the lesion is worrisome. Improvements in machine learning techniques are beneficial for disease diagnosis and image analysis in medicine. Dermatologists can employ automated analysis of these dermatoscopy images to aid diagnosis and decision-making. Deep learning techniques have addressed skin lesion classification (SLC) issues [9]. In contrast to conventional feature extraction methods, using feature selection for skin lesion identification and classification has become increasingly important in recent years. The most crucial step is segmenting the diseased area and choosing essential features from the raw data. Brightness, lesion borders, and color resemblances between the skin and the lesion in skin cancer images complicate and lower classification accuracy [10]. These factors make it necessary to classify skin cancer images in some processes, including image preprocessing techniques, extraction of important features, and classification. The research space identified in the mentioned studies can be

summarized as follows: first, further research is needed to develop precise and efficient methods for detecting and classifying skin lesions. Various techniques have been proposed, such as automatic image processing, deep convolutional neural networks, and ensemble frameworks, but there is still room for improvement in accuracy, sensitivity, and specificity. To ensure their robustness and generalizability, method performance should also be assessed on bigger and more varied datasets. Second, methods for balancing and augmenting data have demonstrated the potential to enhance the classification of skin lesions. More thorough research is nevertheless required to evaluate these technologies' performance across a range of datasets and contexts. Furthermore, research into computationally efficient models is needed to develop reliable for skin cancer detection. Finally, a comparative analysis shows the superiority of ensemble models compared to traditional machine learning algorithms for skin lesion classification tasks, especially in improving performance on large and imbalanced datasets to find out. This study fills this gap by proposing an approach that combines deep learning models and data augmentation for balance techniques to improve performance in skin lesion classification.

This work leverages the advantages of data augmentation to address the imbalance in skin types with a reduced set of images. By generating a well-balanced dataset, we aim to enhance the accuracy of skin classification. The primary contributions of this study are as follows:

- The medical skin lesion images are resized to reduce memory consumption and improve latency, improving latency, thereby enhancing the model's efficiency and applicability in real-world scenarios.
- We employed traditional data enhancement techniques were employed to overcome the imbalanced data between classes, which not only reduce overfitting but also improves robustness of the model.
- An ensemble model based on IncepX-Ensemble was developed specifically to classify seven types of skin lesions based on a weighted average. As a result of this approach, the overall performance of classification is improved by leveraging the strengths of multiple models.
- In order to comprehensively assess the model's performance, we applied a variety of evaluation metrics and a confusion matrix. Through this comprehensive evaluation, a higher standard of performance validation is set in skin lesion classification.

II. RELATED WORK

Researchers have proposed numerous studies in recent years to detect early-stage skin cancer. Most of the latest research is included in this section. According to the World Health Organization (WHO) [11], 13.1 million people are expected to die from cancer worldwide by 2030, most of them in the United States, where skin cancer affects most people. As it penetrates quickly due to abnormal cell growth, skin

134 cancer often spreads to other parts of the human body and 189
135 is a common human disease [12]. Various approaches have 190
136 been tried in recent years. Presents and practices with a 191
137 focus on the health field classify skin lesions. 192

138 Skin cancer's high fatality rates have spurred global 193
139 scientific efforts in its classification and detection. Theoret- 194
140 ical literature on skin cancer is expanding, benefiting from 195
141 ongoing advancements in science and technology. Evolv- 196
142 ing methods for skin cancer classification and diagnostics 197
143 progress from single imaging to targeted classification, em- 198
144 ploying diverse methodologies like computer vision-based 199
145 machine learning. These approaches integrate supervised 200
146 learning and deep learning algorithms for reliable skin lesion 201
147 classification. Notably, the use of Convolutional Neural 202
148 Networks (CNN) has gained prominence, demonstrating 203
149 advanced machine-learning applications in skin cancer clas- 204
150 sification. In the study [13], authors looked into how image 205
151 pre-processing affected how well a saliency segmentation 206
152 method worked for skin lesions. The results confirm that 207
153 the CHC-Otsu approach performs as intended and achieves 208
154 a 92% accuracy rate. Computer-aided diagnostic (CAD) 209
155 [14] systems offer distinct advantages to radiologists to- 210
156 day, with CAD being an integral part of contemporary 211
157 medical imaging. These systems are great because of the 212
158 advanced algorithms for fine-grained detection and clas- 213
159 sification. More recent studies on skin cancer detection, 214
160 [15], [16] have demonstrated that using CAD systems can 215
161 enhance the accuracy of diagnosis and shorten analysis 216
162 time. It is widely used to generate synthetic samples and 217
163 balance datasets using methods such as Synthetic Minority 218
164 Oversampling Technique (SMOTE) and Adaptive Synthetic 219
165 Sampling Approach(ADASYN). Traditional image-based 220
166 data augmentation includes flipping, rotate etc. as well as 221
167 generative models for synthetic medical data generation. 222
168 223

A. CNN METHODS FOR THE CLASSIFICATION OF SKIN LESIONS 224

169 In recent years, machine learning technology based on 226
170 DCNN has been widely established as a cutting-edge tech- 227
171 nology for medical video classification. Some examples of 228
172 such technologies are GoogleNet [17], Microsoft ResNet 228
173 [18], DenseNet [19], MobileNet [20] and EfficientNet [21]. 229
174 Model selection is an area of notable growth. These methods 230
175 eliminate the need for a specific base model by prioritizing 231
176 the selection of unsupervised model types for detection. 232
177 The accuracy of the classified model increased after it was 233
178 trained on a specific dataset. The authors [22] proposed 234
179 RDDNN (Residual Deep Convolutional Neural Network) 235
180 for skin lesion diagnosis and trained and tested six well- 236
181 known skin cancer datasets. The experimental procedure 237
182 begins in two steps. One is to use the original image 238
183 without preprocessing or segmentation, the other is to use 239
184 the segmented image to evaluate network performance under 240
185 different conditions. Hosny et al. [23] proposed a deep 241
186 inherent learning technique to classify seven types of skin 242
187 lesions. Several explanatory strategies were used to validate 243
188

the proposed detailed unique learning. A residual deep convolutional neural network was suggested by the author [24]. Their method primarily modifies the multiple convolution filters for cross-channel correlation and multi-layer feature extraction by shifting the dot product filter rather than the filter along the horizontal axis. In addition to using the ISIC-2019 and ISIC-2020 datasets, the suggested solution solves the issue of imbalanced datasets by converting the picture and label dataset into a vector of images and weights.

In the study by Polat et al. [25], raw skin lesion images were used as input to classify multi-class skin diseases. They claimed that combining a CNN and a one-versus-all approach obtained auspicious results in skin disease classification from dermatoscopy images. Lima et al. [26] proposed combining CNN methods with the Squeeze-and-Excite mechanism in a residual block. The model is trained with the data from scratch for the different classification problems. The findings demonstrate that it is significantly more likely than other ways to have a free and favourable influence. A hybrid-CNN classifier for recognizing numerous skin diseases was proposed by Hasan et al. [27] To produce better depth feature maps of the lesion, Hybrid-CNN fuses three different feature extractor modules. Tang et al. [28] suggested a multi-stage, multi-modal learning system to classify skin lesions with many labels. Two phases make up the suggested model. In the first phase, they construct Level 1 of the fusion, which unifies the feature-level representation of images. They propose a different fusion, training a Support Vector Machine (SVM) cluster with patient meta-data information and multi-label prediction data from the first stage. The patient's meta-data would then be added in the second stage. The proposed network classifies skin lesions with 78.5% accuracy using multiple labels. A two-module structure with lesion segmentation and classification was suggested by Khan et al. [29]. Binary images from the 16-layer CNN model and the improved HDCT structure were trained for lesion segmentation.

B. DATA AUGMENTATION TECHNIQUES

The problem of limited labeled data and class imbalance are due to color space transformations (saturation, contrast, brightness, tone adjustment, etc., of training images), geometric transformations (mirroring, cropping, rotation, translation, etc.), blending images, random deletion, etc. This can effectively increase the size of the training data set [30]. Perez et al. [31] provided a data augmentation scenario for skin lesion classification trained with a CNN classifier. Their study showed that geometric and color space variations can improve classification performance. However, shuffling and randomly deleting images could not able to produce better results. This is because these two data augmentation methods cannot ensure that the transformed image retains the relevant skin lesion features of the original image. Skin lesion classification performance can be improved through data augmentation techniques such as geometric and color space modifications. Since

the distribution of the transformed images is similar to the distribution of the original images and does not add any new visual elements to the training set, there is little performance improvement [32], [33]. Training deep learning models is challenging due to the need for labeled and class-imbalanced datasets. Transfer learning was employed by Qin et al. [32], to enhance classification performance using data augmentation technology based on GANs. To improve the categorization accuracy of cells acquired from bone marrow aspirate smears, Hazra et al. [34] advocated using GAN-based data for the classification purpose.

C. ISIC DATASET DISCUSSION

Jasil et al. [35] used inceptionv3, VGG16, and VGG19 to classify skin lesions, including the transfer learning idea. On the ISIC dataset, they attained accuracy rates of 74% with InceptionV3, 77% with VGG16, and 76% with VGG19. Alhudhaif et al. [36] employed the ISIC dataset to evaluate six transfer learning networks for the classification of skin cancer. Augmentation was applied to images of smaller classes to address imbalances in class sizes within the ISIC dataset. The outcomes revealed that augmentation is impactful for achieving classification success, showing high accuracy and F1 scores with fewer false negatives. Among the transfer learning networks investigated, the Modified DenseNet121 emerged as the most successful model, achieving an accuracy rate of 98.35%. Chatterjee et al. [37] proposed an integrated, computer-aided method that has been employed to identify each of these skin diseases using RFE-based layered structured multiclass image classification. Hosny et al. [38] proposed a method for classifying skin lesions that combined transfer learning with AlexNet that had already been trained. We randomly initialize the weights of the final three replacement layers using the parameters of the original model as beginning values. The most recent publicly available dataset, ISIC 2018, was used to test the suggested approach, which has a 98.7% accuracy rate. An MCML classification technique was proposed by Hameed et al. [39] for a multi-class skin disease categorization. Both sophisticated deep learning techniques and conventional machine learning were employed in the MCML classification algorithm. Using a small dataset, the algorithm demonstrated 96.4% accuracy in testing. They also claimed that the proposed method could significantly enhance classification performances in multi-class skin datasets. A dual attention-based network for skin lesion categorization with auxiliary learning was proposed by Wei et al. [40]. The spatial and channel attention modules are a part of the dual attention system. When trained on the ISIC 2018 and ISIC 2019 datasets, the suggested model has a precision of above 80%. Transfer learning is the process of using a model that has already been trained on huge data, and now we are trying to teach a new model with some new features rather than starting from scratch. Chatterjee et al. [41] proposed an Alzheimer's disease classification mechanism using a voting ensemble approach as a smart

way to investigate the type of disease. A Convolutional Neural Network (CNN) model for the classification of skin diseases was proposed by Kethana et al. [42]. They classified skin illnesses with a 92% accuracy rate using a collection of 10015 images taken from the ISIC 2019 dataset.

D. DATA IMBALANCE CHALLENGES

The problem of imbalanced data has been tackled in a number of ways, especially in the area of object detection when the background (majority) and foreground (minority) classes are not equally represented. The paper by [43], authors proposed accurate automated melanoma classification using dermoscopic images. To solve the problem of interclass similarity, the authors of this article developed a unique, lightweight DeepCNN-based framework to distil information in the context of melanoma classification. In the study, they proposed to apply cost-sensitive learning with loss of focus to solve the critical class imbalance problem and achieve higher sensitivity values.

To ensure accurate predictions, Akilandasonmya et al. [44] using the ResNet50 method (SCSO-ResNet50), used a novel approach called sand cat swarm optimization to distinguish between known features and deeply hidden features. Apply enhanced harmony search (EHS) techniques to minimize data dimensions and maximize properties. Ensemble classifiers can help detect cancer early. This research proposes a method to enhance the DCNN classification performance on imbalanced datasets for multi-class medical images. The authors [45] suggest utilizing optimal domination of weights across asymmetric classes with majority and minority categories to solve the issue of imbalanced samples, as well as a robust feature extraction strategy for improved training and loss function optimization.

E. ENSEMBLE LEARNING BASED

In recent years, the complexity and drawbacks of real-world problems have increased, requiring the development of new advance algorithms, transfer learning techniques, meta-leaning models (ensembles), and more reliable models. The authors [46] used a meta-learning algorithm with a stacking ensemble method to achieve better performance. Using a weighted average ensemble learning model that integrated multiple transfer learning models, Rahman et al. [47] identified 18,730 lesion pictures from the ISIC 2019 and HAM10000 datasets. The ensemble model with a weighted mean achieved 94%, whereas the ensemble model with a basic mean achieved 93%. One of the models proposed by the study authors is a multiscale deep ensemble [48]. To extract more features, the proposed model precisely cuts out the lesion area. They then linked the cropped image's features to several metrics to classify the results. As a result, HAM10000 fragments of the training dataset are reported to outperform other approaches.

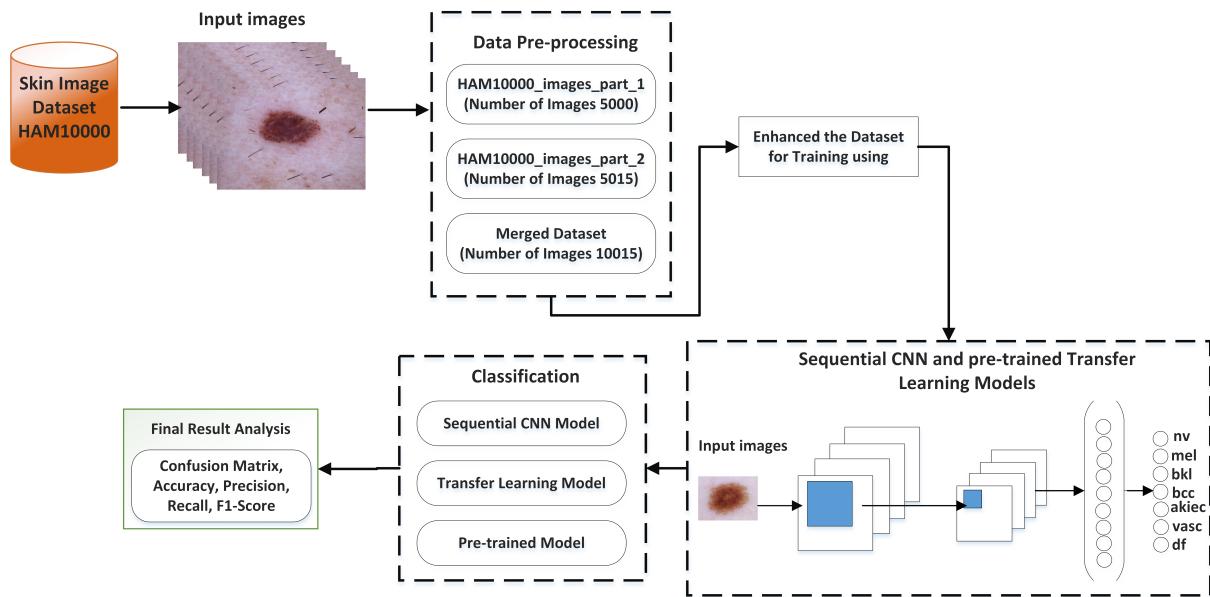


FIGURE 1. Flowchart of the proposed multi-class skin lesion classification.

III. METHODOLOGY

This work presents a multiclass skin cancer classification methodology using a pre-trained transfer learning model with best-selected features. Fig. 1 provides a complete flowchart of the methodology. Additionally, the study's data collection and the image preprocessing techniques applied to it. The proposed IncepX-Ensemble model utilized for the classification problem is briefly presented, and data augmentation was done to balance the data set. We integrated the inceptionV3 model with the Xception model, which effectively made the ensemble model improve performance and computation efficiency. In this research, we introduce an advanced model that leverages weighted average ensemble learning to classify seven different types of skin lesions. The model combines the strengths of two well-known deep neural networks, InceptionV3 and Xception, as its base components. InceptionV3 is known for its efficiency in handling complex visual tasks through its unique architectural features. Xception, on the other hand, employs depthwise separable convolutions to enhance performance and accuracy. By integrating these two networks, our ensemble model named the IncepX-Ensemble model aims to improve classification accuracy. This approach allows us to effectively utilize the diverse capabilities of both networks to achieve superior results in skin lesion classification. The dataset was improved using a data enhancement technique after the preprocessing stage. The proposed model was applied to both the unbalanced original dataset and the datasets that had been balanced using balancing procedures. Following the training, experiments for testing and validation were conducted along with the classification procedure.

A. DATA ANALYSIS

Dermatoscopy can improve the doctor's view of the skin, helping to diagnose different skin conditions and consider suspicious skin lesions. Compared to ocular inspection, this approach is frequently utilized to improve the identification of benign and malignant pigmented skin lesions [49]. One of the most extensive datasets of 10015 dermoscopic images gathered over 20 years is the HAM10000 (Human Against Machine) dataset. The information was collected from research labs at several universities, the dermatology department at the Medical University of Vienna in Austria, and Cliff Rosendahl's skin cancer clinic in Queensland, Australia [50]. All records of the HAM10000 data set are archived in the Harvard Dataverse. A total of 10,015 skin lesion images from seven main kinds of skin lesions make up the HAM10000 dataset. Fig. 2 displays the sample images for each skin lesion type. Fig. 3 exhibits the distribution of samples across different classes in the original dataset using a bar representation. The distribution of classes on the data set is disrupted by the imbalanced number of samples belonging to the Melanocytic nevi (nv) class, as seen in Table 1. A high-class imbalance exists in the dataset, with more than two-thirds of the images falling under the nv class.

Table 1 provides a detailed breakdown of the distribution of image samples across the HAM10000 dataset for the most crucial diagnostic classes. The data set includes 6705 images of melanocyte nevi (nv), 1113 images of melanoma (mel), 1099 images of benign keratosis-like lesions (bkl), 514 images of basal cell carcinoma (bcc), 327 images of actinic keratosis (akiec), and 142 images of Vascular lesions (vasc) images and 115 dermatofibroma (df) images.



FIGURE 2. Sample of each skin type are shown in the figure.

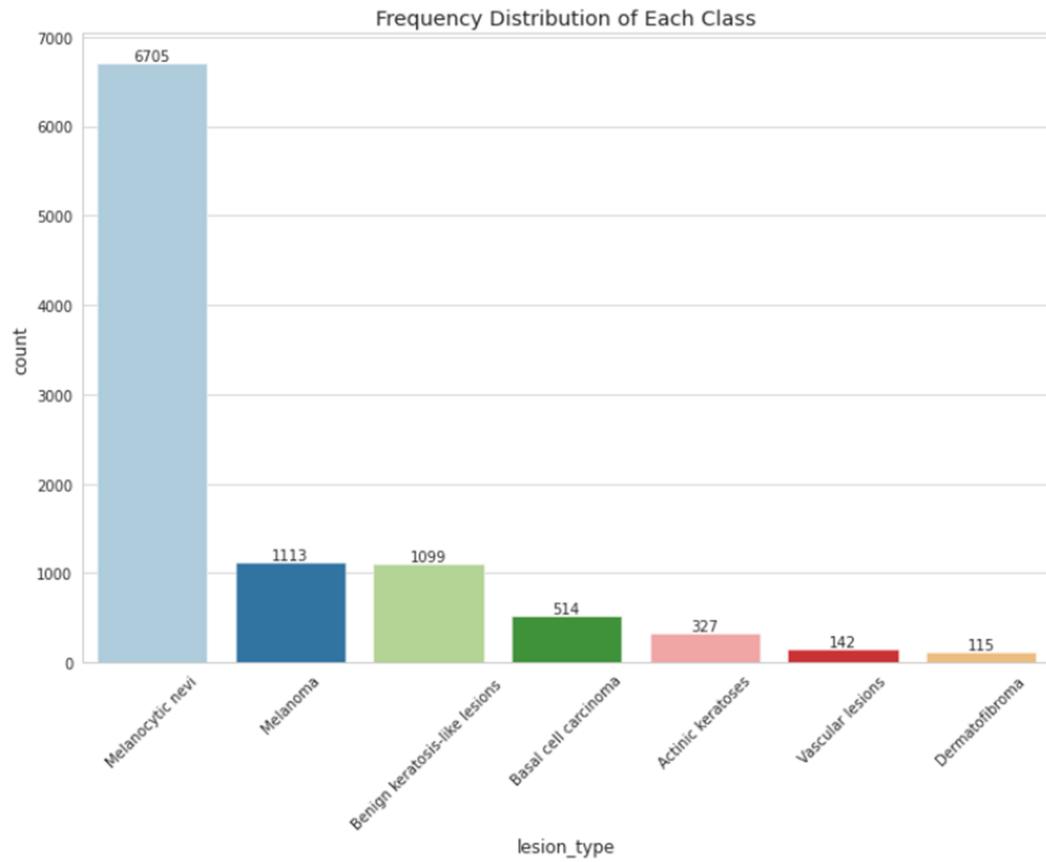


FIGURE 3. Number of instances distribution across the seven classes is shown in the figure.

TABLE 1. HAM10000 dataset distribution of the images among the seven main skin lesions classes.

Class Name	Abbreviate Diagnostic Category	No. of Images	Percentage
Melanocytic nevi	nv	6705	66.95%
Melanoma	mel	1113	11.11%
Benign keratosis-like lesions	bkl	1099	10.97%
Basal cell carcinoma	bcc	514	5.13%
Actinic keratoses	akiec	327	3.27%
Vascular lesions	vasc	142	1.42%
Dermatofibroma	df	115	1.15%

However, to effectively train a deep learning-based model, we need sufficient balanced data. Intentionally generating the necessary data helps achieve data balancing, preventing biased model training. This precaution is taken when the imbalanced data could lead to ongoing bias in model training, particularly favouring classes with a substantial number of samples. In the pie chart diagrams, the distribution of diagnosis categories in each class is depicted in the left

side pie chart, which shows in Fig. 4. On the right side of Fig. 4 shows the distribution of the diagnosis method. More than 50% of the lesions in the images were confirmed histopathologically, and the remaining cases were confirmed by follow-up examinations, expert consensus, or in vivo confocal microscopy. Each lesion's diagnosis was verified using a particular method. A total of 5,340 (53.3%) lesions were confirmed histopathologically, 3,704 (37.0%) by

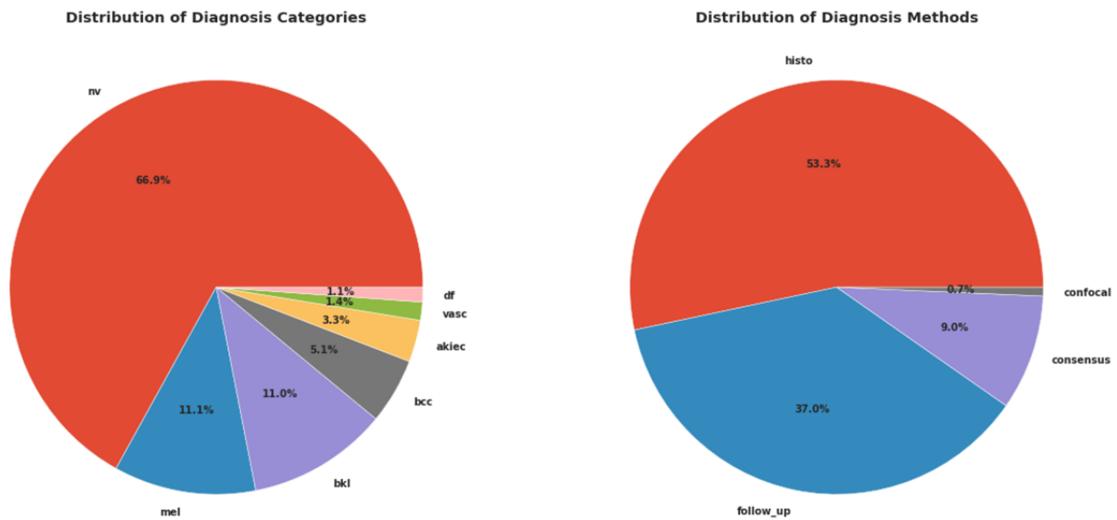


FIGURE 4. Pie charts representing the distribution of skin lesion data. As illustrated in the left graph, different diagnosis categories (e.g., nv, bkl, df, vasc) are distributed by percentage, while different diagnosis methods (e.g., histo, follow-up) are shown in the right graph.

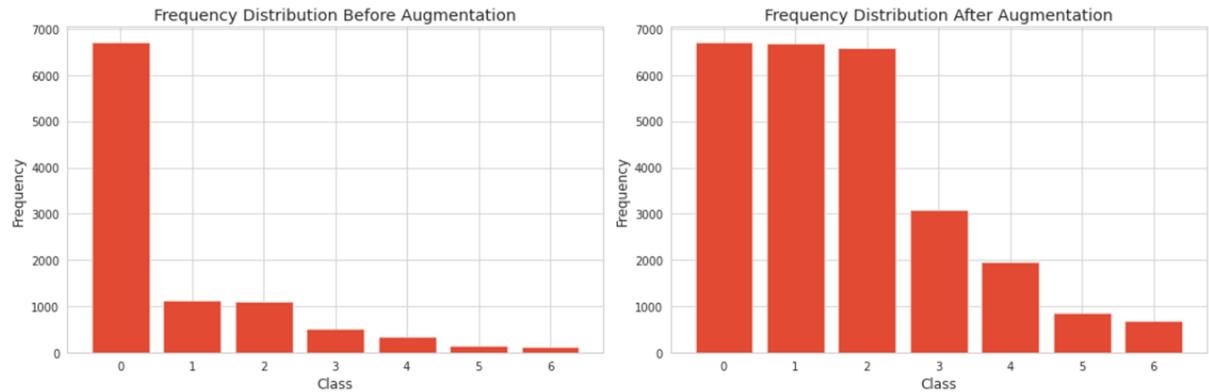


FIGURE 5. The figure depicts the frequency distribution of classes before and after data augmentation. The diagram on the left shows the frequency distribution of the original class labels before data expansion. The graph on the right shows the frequency distribution after applying data augmentation techniques.

429 follow-up, 902 (9.0%) by expert consensus, and 69 (0.7%)
430 were confirmed based on ground truth.

431 1) Image Preprocessing

432 Each image in the dataset has a dimension of 600x450
433 pixels. We expanded the dataset by using rotation and
434 flipping as part of conventional image-enhancing methods.
435 We artificially increased the size by rotating the dataset
436 and flipping it horizontally and vertically. Because neural
437 networks need a lot of labelled data to train, dataset size has
438 typically been a problem in the medical field. To overcome
439 the data imbalanced issue, we balanced the dataset by using
440 data augmentation to train the model with balance. Data
441 augmentation is helpful when a significant data imbalance is
442 distributed among the dataset's classes. Data augmentation
443 may impact model training if the data is unbalanced or
444 very small. This can lessen the deep neural network's
445 overfitting. To address the problem of data imbalance in
446 most of the classes in the HAM10000 dataset, we applied
465

448 data augmentation techniques.

449 The precise kind of data augmentation method employed
450 in the investigation is described in depth. A variety of data
451 augmentation methods, such as rotation and inversion, were
452 used in our methodology. I specifically reversed the image
453 both horizontally and vertically and rotated it 90 degrees
454 clockwise, 90 degrees counterclockwise, and 180 degrees.
455 These techniques help increase the size of the training
456 data set, preventing overfitting and exposing the model
457 to different transformations of the same image, thereby
458 increasing the robustness of the model. Our methodology
459 included several key steps. First, the images are read from
460 the dataset. I then created a new image by rotating and
461 flipping using the "producer_new_img" function. For im-
462 ages from classes 1 to 6, additional augmented images were
463 created to balance the class distribution. Next, we extracted
464 the labels corresponding to the images and assigned them
465 to the augmented images as well. To illustrate the impact
466 of the enhancement technique, Fig. 5 shows the frequency

466 distribution of classes before and after enhancement.

521

467 **B. PROPOSED MODEL**

523

468 A powerful technique to accelerate model development is 524
469 transfer learning in the field of deep learning. Compared to 525
470 creating a model from scratch, transfer learning drastically 526
471 cuts down on computational resources and training time. 527
472 Moreover, it can be very helpful when dealing with small 528
473 datasets. To train the model, the image classification field 529
474 often needs a sizable collection of tagged images. But by 530
475 using the advantage of transfer learning, we may "fine- 531
476 tune" a pre-trained model trained on ImgaeNet. Compared 532
477 to building a model from scratch, this method uses less time, 533
478 which makes it a useful tool for applications requiring small 534
479 datasets. Transfer learning techniques are utilized to enhance 535
480 the performance of machine learning algorithms using la- 536
481 beled data. These efforts involve learning and applying one 537
482 or more source tasks to improve learning in related fields. 538
483 Transfer learning has been studied as a machine learning 539
484 process to address various problems. It includes pre-training 540
485 models on large datasets and subsequently retraining them 541
486 at different levels on smaller training sets. Adjustments to 542
487 the initial layer of the pre-trained network can be made if 543
488 necessary. Fine-tuning the parameters of the model's final 544
489 layer allows it to learn from new datasets. According to the 545
490 new task, pre-trained models are retrained with a smaller 546
491 new dataset, modifying the model weights accordingly. This 547
492 means that newly developed neural network parameters are 548
493 not built from scratch. Although deep learning algorithms 549
494 can achieve higher functionality or performance for many 550
495 problems, they require extensive data and training time. 551
496 Thus, reusing pre-trained models for similar tasks can be 552
497 highly beneficial. In our approach, we used two pre-trained 553
498 models: InceptionV3 and Xception. These models, initially 554
499 pre-trained on the ImageNet [52] dataset, were fine-tuned 555
500 using the skin lesion dataset. 556

501 The ImageNet dataset, comprising millions of images, 557
502 was used to train the Inception [53] and Xception [54] 558
503 network models. The InceptionV3 network, with 48 layers, 559
504 and the Xception network, with 71 layers, both require input 560
505 images of size $299 \times 299 \times 3$ pixels. Fig. 6 and Fig. 7 561
506 illustrate the structural layouts of InceptionV3 and Xception, 562
507 respectively. 563

508 We developed a special image resizing method that takes 564
509 into account the different sizes of the input images while 565
510 preserving important features. The image size was reduced 566
511 to 299×299 pixels to retain important visual information 567
512 while maintaining computational efficiency. After much 568
513 testing, we chose this dimension so that resized images 569
514 retain a high level of information necessary for accurate 570
515 classification. 571

516 We also used data augmentation methods such as flipping 572
517 and rotating to add variety and creativity to our training 573
518 set. By mimicking the variance found in the real world 574
519 and avoiding overfitting, this extension helps improve the 575
520 generalization of the model. We use these changes to 576

improve the model's detection and classification efficiency
for new instances by ensuring that the resized images fit the
model's input size while preserving their original properties
and variants.

InceptionV3 addresses common issues such as congestion
and yield by employing asymmetric filters and bottlenecks,
and replacing larger filters with smaller ones, leading to ef-
fective outcomes. This architecture optimizes the network's
performance by managing computational complexity and
enhancing feature extraction efficiency. The design includes
factorized convolutions and aggressive regularization, which
contribute to its robustness in handling complex visual
tasks. On the other hand, Xception is designed to be more
efficient and less complex. It achieves more precise and ef-
ficient results by separately applying cross-channel and spatial
correlations, thus improving the model's accuracy and speed.
Additionally, the Xception model utilizes depth-wise
separable convolution and employs cardinality to develop
more effective abstractions, enhancing its performance. This
separation allows for the creation of more nuanced represen-
tations, making Xception particularly effective for various
image classification tasks. By leveraging these advanced
architectures, both InceptionV3 and Xception achieve high
accuracy on large-scale image classification benchmarks,
demonstrating their suitability for diverse applications in
computer vision.

InceptionV3 and Xception have been successfully applied
in various research areas, including object detection, image
segmentation, and medical image analysis. Their high per-
formance on the ImageNet dataset has led to widespread
adoption in tasks requiring high accuracy and robustness.
InceptionV3's factorized convolutions and aggressive regu-
larization make it well-suited for handling complex visual
tasks while managing computational complexity. On the
other hand, due to using depth-wise separable convolutions
both efficiency and performance are improved by Xception
-leading to accomplished results for various standard image
classification tasks. Both of these models benefit from being
pre-trained on large amounts of data on ImageNet which
makes them a great option for transfer learning in low-data
domains. These models can be computationally challenging,
though, which makes it demanding to train or draw con-
clusions from them. Their architectures are intricate, and
it can be challenging to reuse or fine-tune a pre-trained
model without considering the unique limitations of your
intended application. We decided on these models because
their modern architectures and state-of-the-art performance
on large-scale image classification tasks prepare them well
for our task of increasing the accuracy of skin lesion
classification. By combining the strengths of the proposed
IncepX-Ensemble modes by sharing the free features of
Xception and InceptionV3 and maximizing their benefits.

First, we inserted pre-trained ImageNet weights into the
InceptionV3 and Xception models, leaving out the top layer.
After that, we concatenated each model's output and added
a global spatial average pooling layer to create a composite

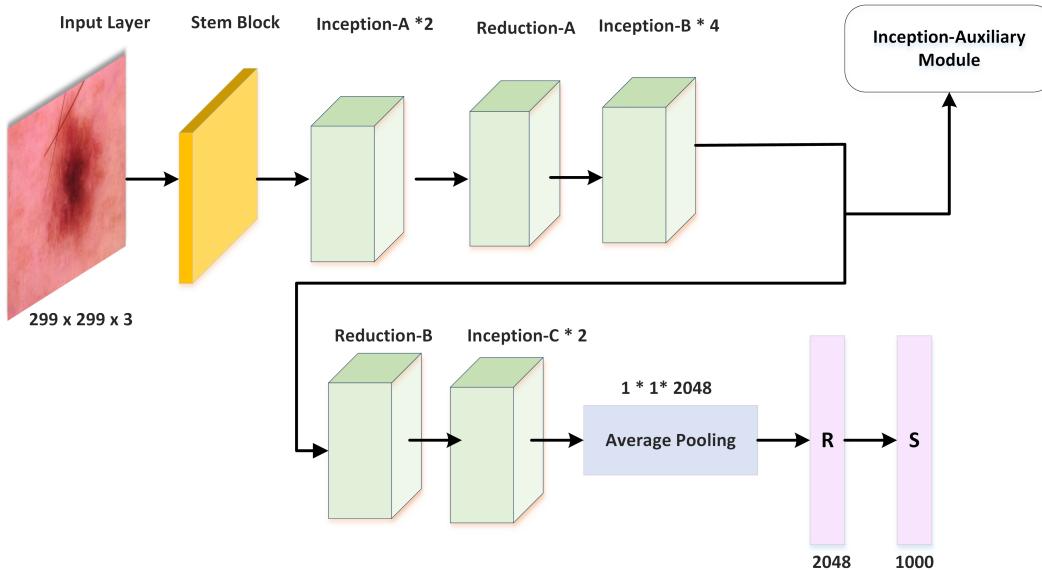


FIGURE 6. Network architecture of InceptionV3.

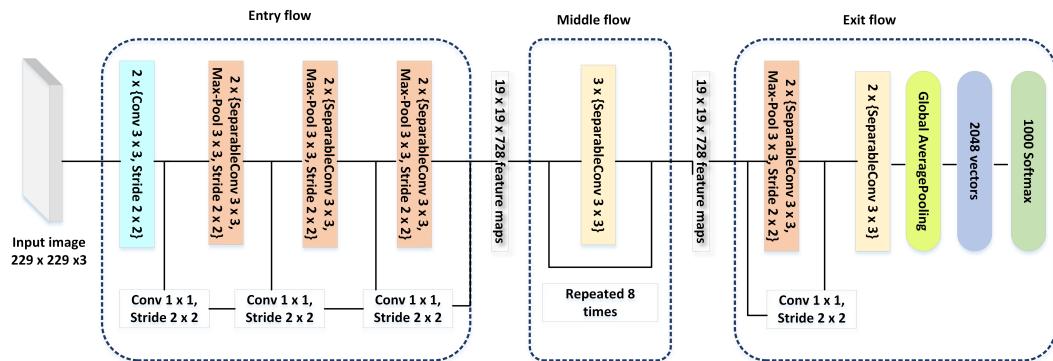


FIGURE 7. Network architecture of Xception [51].

feature vector. To create the final prediction, a softmax layer with seven classes was added after a dense layer with 1024 neurons and ReLU activations. To increase its capacity for generalization, this ensemble model was assembled using the Adam optimizer and trained utilizing data augmentation methods.

583 1) Ensemble Learning

Using ensemble learning, multiple models can be combined to improve performance and computation efficiency. Ensembles of deep neural networks always perform better than single models, regardless of the number of models. Each model was weighted equally in this study, using average ensemble learning.

$$P = \frac{\sum A_i}{B} \quad (1)$$

The probability of model i can be calculated from equation 1, and the total number of models is B .

Deep learning models often yield different results due to their varying architectures and levels of complexity. By weighting supervised learning models based on their performance, we can ensure that each model's output is optimized to the fullest extent possible. However, identifying the optimal combination of model weights is a challenging task. To address this issue, we utilized the grid search technique, as shown in Fig. 8. We explored a total of 2000 different weight combinations. The search process continued until all possible combinations were evaluated. Ultimately, this method gave us the optimal weights, maximizing our evaluation metrics.

In unsupervised learning environments where the optimal outcome is unknown, clear tactics must be implemented. Weights can be assigned according to the consistency and agreement between models using methods such as mutual information and consensus clustering. In addition, the self-learning technique of the pseudo-labeler serves as an alternative to performance evaluation and supports weighting

procedures in the event of missing labeled data.

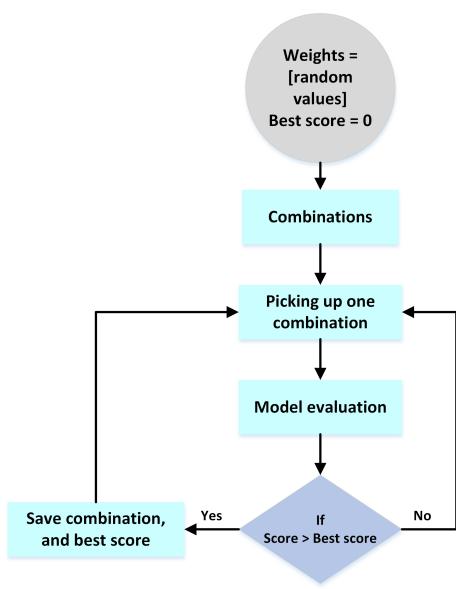


FIGURE 8. Weights are determined by grid search.

Taking advantage of ensemble learning technique to improve the performance and robustness of machine learning models, is the main contribution of this study. Combining several models helps us take advantage of their strengths and minimize their weaknesses. This approach often provides better performance than a single model can achieve. Specifically, in this study, we used an average ensemble method in which each model contributes equally to the final prediction. Techniques for grid search are very useful since they make it possible to explore the hyperparameter space systematically. Finding the ideal weighting combination is computationally demanding, but the results are well worth the effort. Gives the best-performing models more weight, enabling the ensemble model to utilize each model to its fullest potential. The grid search method for determining the ideal model weights is depicted in Figure 8. We rigorously assess every conceivable combination, starting with random values, to identify the optimal combination. In conclusion, grid search methods in conjunction with ensemble learning strategies offer a potent means of maximizing model performance. By doing so, extremely effective ensemble models are produced, maximizing the benefits of each individual model.

IV. EXPERIMENTAL RESULTS ANALYSIS

The detailed outcomes of the experiments are discussed, and a performance analysis of the results is presented in this section. A description of the evaluation metrics and details of system implementation is provided. In this section, the experimental outcomes are deployed with the help of tables, confusion matrices, and accuracy-loss plots. All machine learning models were trained and tested on a Ubuntu system,

with GEFORCE RTX 4080 GPU, and 32GB of RAM. The environment is having Python version 3.7.7.

A. EVALUATION METRICS FOR PERFORMANCE ANALYSIS

The suggested model is trained over 100 epochs using a batch size of 32 and Adam as the optimizer. The model is trained using an initial learning rate 0.0001 while efficiently using the number and kernel sizes. We chose this learning rate based on preliminary experiments and existing literature suggesting that a lower learning rate can improve training stability. Nevertheless, we acknowledge the risk of training slowing down if this low learning rate is used initially. Table 2 summarises the model hyperparameters in general. In the model training, the activation function ReLU is utilized. We split the total dataset 80% as training data and 20% as testing data.

Calculated metrics such as accuracy, precision, recall, and f1-score are used to assess how well the trained model performed with test samples. Images of skin lesions without enhancement serve as testing samples. Knowing the model's performance for hypothetical samples is the goal here. The metrics stated in Equations 2–5 were used to assess how well each system performed when classifying images of skin lesions for the HAM10000 dataset.

$$\text{Accuracy}(acc) = \frac{T_P}{T_P + T_N + F_P + F_N} \quad (2)$$

$$\text{Precision}(pre) = \frac{T_P}{T_P + F_P} \quad (3)$$

$$\text{Recall}(rec) = \frac{T_P}{T_P + F_N} \quad (4)$$

$$F1 - \text{score} = \frac{2 * T_P}{2 * T_P + F_P + F_N} \quad (5)$$

These formulas include true positive T_P and true negative T_N , representing various correctly detected positive and negative samples. Additionally, the terms false positive F_P and false negative F_N describe the quantity of incorrectly detected positive and negative samples.

Four comparisons were made to demonstrate the proposed model's effectiveness. First, we evaluated the classification performance using the CNN models with the IncepX-Ensemble model when data augmentation was used. The significance of the proposed model was thoroughly assessed through a comprehensive evaluation that included comparisons with and without data augmentation, as highlighted in Tables 3 and 4. It is important to emphasize that the results obtained indicate a notable superiority of the proposed model over alternative counterparts in terms of accuracy, specificity, and sensitivity. This heightened performance underscores the robustness and effectiveness of our model in addressing the challenges posed by the dataset. Data augmentation further enhances the model's ability to generalize

TABLE 2. Model hyperparameters summary.

Activation	Optimizer	Batch Size	Epochs	Learning rate	Loss function
ReLU	Adam	32	100	0.0001	categorical_crossentropy

TABLE 3. Comparison of the different model performances with original imbalanced dataset.

Model	Accuracy	Precision	Recall	F1-score
CNN	0.82	0.81	0.81	0.83
Resnet50	0.83	0.83	0.82	0.83
InceptionV3	0.85	0.85	0.83	0.83
proposed model	0.86	0.86	0.87	0.86

TABLE 4. Comparison of the different model performances with data augmentation.

Model	Accuracy	Precision	Recall	F1-score
Data augmentation with CNN	0.919	0.919	0.921	0.922
Data augmentation with Resnet50	0.935	0.935	0.933	0.935
Data augmentation with InceptionV3	0.957	0.958	0.953	0.959
Data augmentation with proposed model	0.988	0.988	0.989	0.989

TABLE 5. Comparison of proposed model performance using HAM1000 dataset with some existing models.

Model	Dataset	Accuracy
Darknet-53 [55]	HAM10000	95.8%
EfficientNet [56]	HAM10000	84.3%
MobileNet + LSTM [57]	HAM10000	85.3%
DenseNet201 network [58]	HAM10000	92.8%
DenseNet169 [59]	HAM10000	91.1%
VGGNET-16 [60]	HAM10000	85.6%
Proposed Model	HAM10000	98.8%

and perform well across diverse scenarios, contributing to its superiority over alternative approaches.

Computation:

• **Sequential CNN Model:** Sequential CNN model: This model includes standard convolution and pooling layers, resulting in a relatively simple calculation process. The computational complexity of a single CNN is generally $O(n \times m \times k \times d)$, where n is the number of filters, m is the filter size, k is the number of input channels, and d is the network depth.

• **IncepX-Ensemble Model:** Ensemble models combine multiple base models to increase the computational load through additional forward and backward propagation during training. The combined complexity can be approximated by $O(n \times m \times k \times d \times L)$. Here L is the number of models in the set. However, parallel processing functions reduce the overall computational effort and enable efficient training.

B. COMPARISON WITH CONFUSION MATRIX

A confusion matrix is a table commonly used to describe the performance of classification models. Compare the actual target value with the model predicted value.

• **Number of diagonals:** The confusion matrix's diagonal elements show the instances in which the actual and expected labels coincide. The amount of accurate predictions for each class is represented by this value.

The number at position (i, i) , for instance, shows how many times the model accurately predicted class i . When the diagonal value is high, it indicates that the model is successful in accurately classifying instances of a given class.

• **Numbers on the Off-Diagonal:** Off-diagonal elements represent cases where the predicted label does not match the actual label. This value indicates the number of incorrect predictions. For example, the number at the (i,j) position (where $i \neq j$) indicates how many times the model incorrectly predicted class j even though the actual class was i . A lower value on the off-diagonal line is preferable, as it indicates less misclassification.

We can learn more about the errors that occur in our model by looking at the confusion matrix. For instance, a high value for a specific off-diagonal line suggests that the model frequently mixes together different classes. This data aids in identifying particular model flaws and serves as a roadmap for future enhancements.

A confusion matrix is invaluable in evaluating the performance of a multi-class image classification model, providing a detailed breakdown of the model's success in classifying instances for each category. Fig. 9 shows the class-wise predictions using the confusion matrices. On the left is the predicted outcome based on the CNN model, and on the right is the predicted outcome based on IncepX-Ensemble.

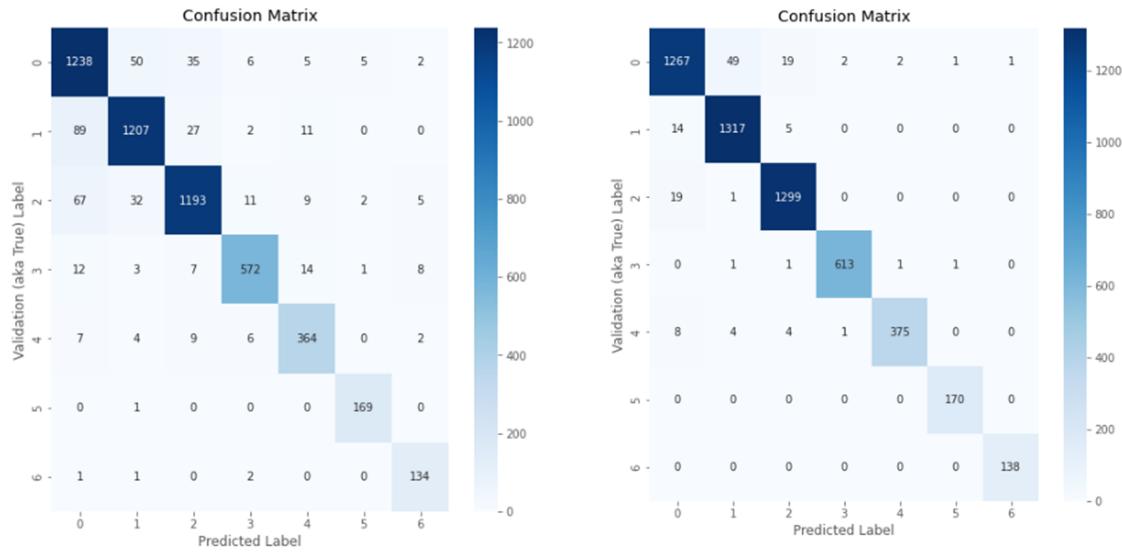


FIGURE 9. Confusion matrices showing the results of skin lesion classification. Left: Results from the CNN model predictions. Right: Results from the proposed IncepX-Ensemble model predictions.

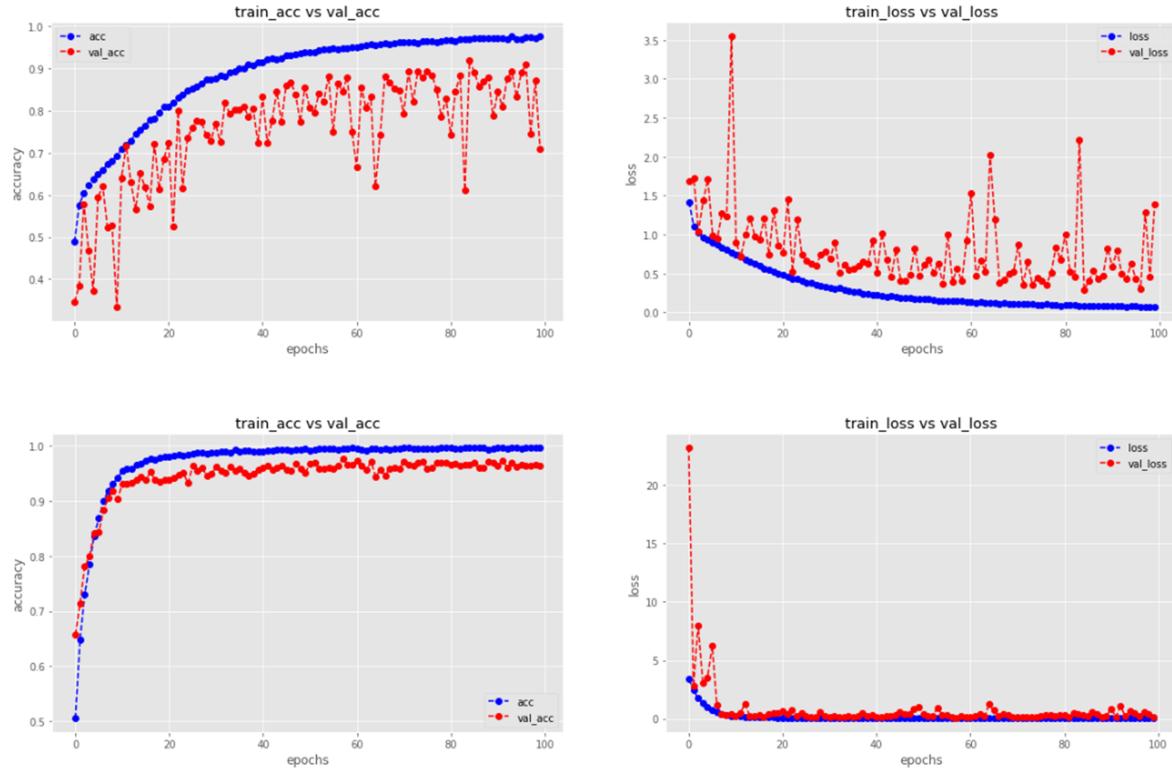


FIGURE 10. Training and validation loss and accuracy results over time for both models. The top row shows the learning curves for the CNN model, while the bottom row shows the learning curves for the IncepX-Ensemble model.

In the confusion matrices, label 0 represents melanocytic nevi (nv), label 1 represents Melanoma (mel), label 2 represents Benign keratosis-like lesions (bkl), label 3 represents Basal cell carcinoma (bcc), label 4 represents Actinic keratoses (akiec), label 5 represents Vascular lesions (vasc) and label 6 represents Dermatofibroma (df). The predicted classification result with the CNN model, the number of misclassified images for each class were, class 0 with 103

misclassified images for each class were, class 0 with 103 misclassifications, class 1 with 129, class 2 with 126, class 3 with 45, class 4 with 28, class 5 with 1, class 6 with 4. The proposed IncepX-Ensemble model improved these results: class 0 had 74 misclassifications, class 1 had 19, class 2 had 20, class 3 had 4, class 4 had 17, and classes 5 and 6 achieved 100% accuracy.

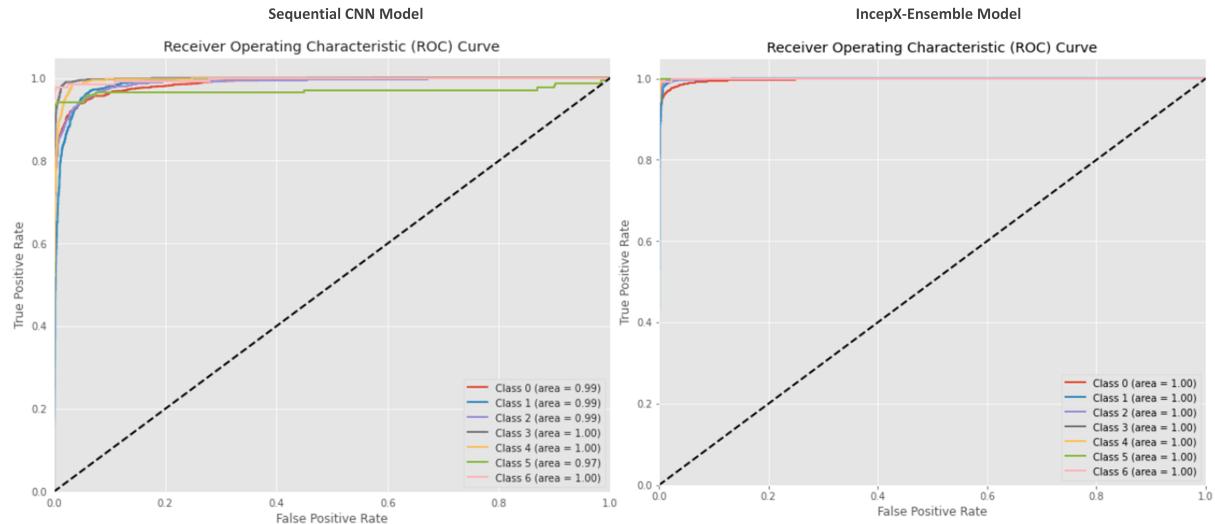


FIGURE 11. Receiver Operating Characteristic (ROC) curves comparing the performance of the Sequential CNN model (left) and the IncepX-Ensemble model (right) on the skin lesion classification task.

Table 5 presents a comparative analysis of the proposed model's performance accuracy with several existing models using the HAM10000 dataset. The table outlines each model, including Darknet-53, EfficientNet, MobileNet + LSTM, DenseNet201 network, DenseNet169, and VGGNET-16, along with their corresponding accuracies. Notably, the proposed model achieves an impressive accuracy of 98.8% on the HAM10000 dataset, surpassing the performance of all other models listed. This considerable accuracy underscores the efficacy of the proposed model in accurately classifying skin lesions. Furthermore, it is worth highlighting that the proposed model achieves this high level of accuracy while maintaining a minimal number of trainable parameters and requiring a short calculation time. Such efficiency and effectiveness make the suggested model a promising candidate for skin lesion classification tasks, demonstrating its potential to contribute to medical image analysis and diagnosis advancements.

TABLE 6. Training Time Comparison.

Model	Training Time (seconds)
Sequential CNN	12,900
IncepX-Ensemble	12,100

We have recorded over 100 generations of training time spent on both the Sequential CNN model and the proposed IncepX ensemble model. The results are summarized in Table 6. Training the sequential CNN model took 12,900 seconds and training the IncepX ensemble model took 12,100 seconds. We found that the IncepX ensemble model had slightly faster training times despite the added complexity of combining multiple models. Thanks to the efficient parallelism capabilities of ensemble approaches, multiple models can be trained simultaneously to optimize overall calculation time.

Fig. 10 displays training and validation data's accuracy and loss curves. It shows the training history of the CNN and the IncepX-Ensemble model with 100 epochs. This is a crucial graphic to ensure the model gets better and more accurate with each passing epoch as it strives to maximize the goal function. The training accuracy and loss graph plot for the sequential CNN model is shown in the figures above, while the accuracy and loss graph plot for the IncepX-Ensemble model is shown in the graph plot below. The accuracy and loss of a sequential CNN model fluctuate during training. However, in the case of the proposed model, they increase steadily up to the 17th epoch, after which they increase at a moderate rate, and then the growth decreases until the 80th epoch. The loss also declines at roughly the same rates.

Table 7 provides a comprehensive breakdown of each class's precision, recall, and F1-score metrics in both the Sequential CNN model and the proposed IncepX-Ensemble model. These metrics offer insights into the models' classification performance across various skin lesion types. Upon analysis, it is evident that the proposed IncepX-Ensemble model performs better than the Sequential CNN model. For instance, in the class of dermatofibroma, the proposed ensemble model achieves a perfect precision, recall, and F1-score of 1.00, indicating its ability to classify instances of dermatofibroma with precision and sensitivity. Similarly, the class of vascular lesions also attains a perfect classification performance with the proposed ensemble model, achieving precision, recall, and an F1-score of 0.99 or higher. These results highlight the robustness and effectiveness of the proposed model in accurately identifying these specific skin lesion types. Overall, the classification performance of the proposed IncepX-Ensemble model consistently outperforms that of the Sequential CNN model across all classes, reflecting its enhanced capability in discerning subtle distinctions

TABLE 7. A class-wise classification report comparing the Sequential CNN model and IncepX-Ensemble.

Class Name	Sequential CNN model				IncepX-Ensemble Model			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
Melanocytic nevi-0	0.89	0.92	0.90	1341	0.97	0.94	0.96	1341
Melanoma-1	0.93	0.90	0.92	1336	0.96	0.99	0.97	1336
Benign keratosis-like lesions-2	0.94	0.90	0.92	1319	0.98	0.98	0.98	1319
Basal cell carcinoma-3	0.95	0.93	0.94	617	1.00	0.99	0.99	617
Actinic keratoses-4	0.90	0.93	0.92	392	0.99	0.96	0.97	392
Vascular lesions-5	0.95	0.99	0.97	170	0.99	1.00	0.99	170
Dermatofibroma-6	0.89	0.97	0.93	138	0.99	1.00	1.00	138
Macro avg	0.92	0.94	0.93	5313	0.98	0.98	0.98	5313
Weighted avg	0.92	0.92	0.92	5313	0.98	0.98	0.98	5313

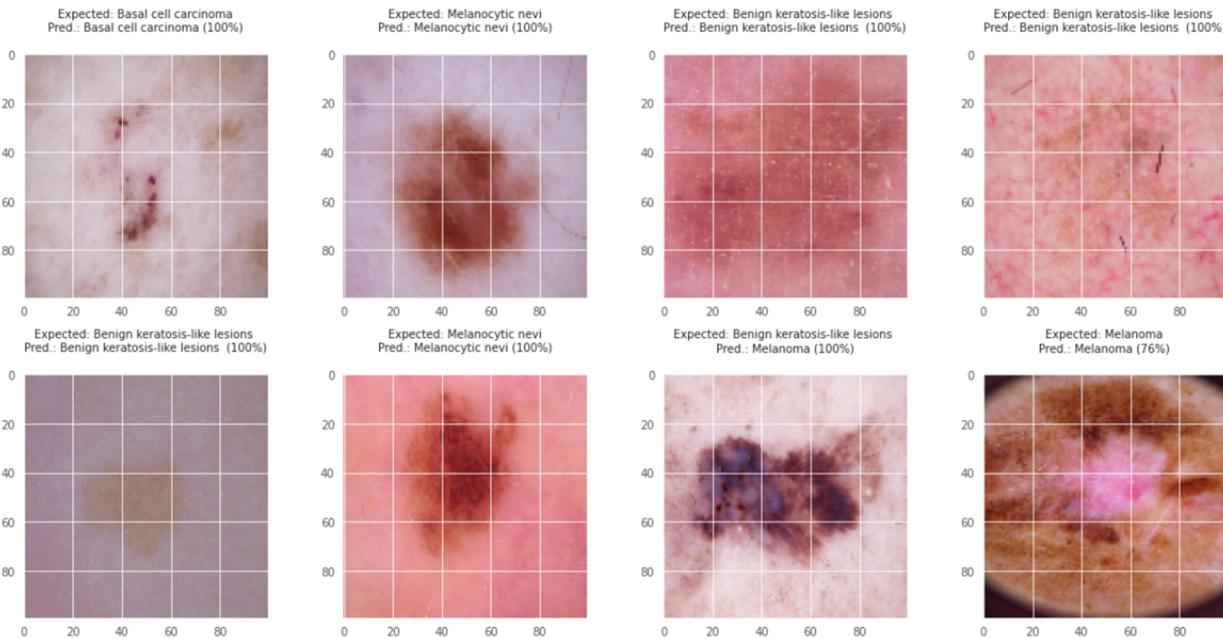


FIGURE 12. Samples of misclassification in skin lesion classification. The image depicts instances where the model's predictions diverge from the actual skin lesion types.

among different skin lesion categories. This suggests that the proposed IncepX-Ensemble model holds promise for improving diagnostic accuracy and reliability in skin lesion classification tasks.

In Fig. 11, the ROC curves shows the performance of two models, tried on skin lesion classification: The Sequential CNN model and the IncepX-Ensemble model. In both the left and right ROC curves, the Sequential CNN model (left ROC curve) exhibits high performance with areas under the curve (AUC) around 0.99 to 1.00 for each class, indicating that the model is remarkably sensitive and specific. As a comparison, the IncepX-Ensemble model (right ROC curve) achieves perfect classification for all classes, with an AUC of 1.00. This concludes that the IncepX-Ensemble model is more accurate and reliable in performing this particular classification task. In summary, both of the models are performing so well, but the IncepX-Ensemble model has perfect separation between true-positive and false-positive rates which will make it more reliable in all classes for skin lesion classification.

Fig. 12 shows a few samples of misclassification or incorrect classification. The predicted result of the model is shown in the figure. An expected (original label of the class) and predicted (recognised by the model) class with an accuracy value is shown at the top of each test image. The percentage indicates how accurately the skin images have been classified. As depicted in the figure, third image from the top left corner, the original is Benign keratosis-like lesions, but the model had misclassified as Melanocytic nevi with a confidence score of 76%.

Fig. 13 compares skin lesion classification results with test images. As shown in the figure, the proposed model correctly predicted most of the test samples. An expected (original label of the class) and predicted (recognised by the model) class with an accuracy value is shown at the top of each test image. For example, from the top left corner, the actual type is benign keratosis-like lesions, and the prediction class is also composed of benign keratosis-like lesions, with a value of 100%. The percentage represents the classification result of the proposed model with a confidence

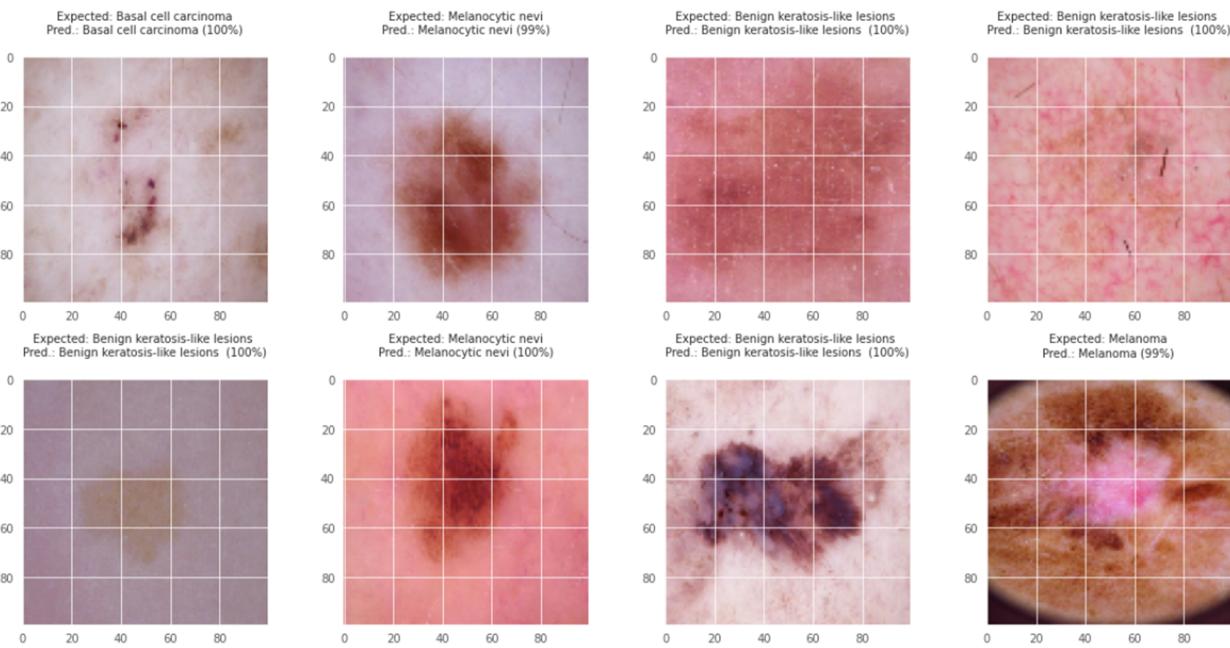


FIGURE 13. Comparative analysis of skin lesion classification results. The image showcases the model's precise predictions for test images, with each labeled to display both the expected and predicted classes.

rate of 100%. In this example, the percentage represents the classification result produced by the proposed model, indicating a high level of confidence in the prediction.

V. CONCLUSION

For skin lesion classification using dermoscopy images, an ensemble learning model based on deep learning is proposed. Inaccurate diagnoses and various skin lesions contributed to the patient's death. One of the deadliest illnesses, melanoma, can be fatal if discovered too late. To detect skin lesions early, doctors have developed many methods based on deep learning. As a result of this study, we were able to achieve satisfactory results in diagnosing skin lesions using ensemble techniques. In this experiment, we propose a transfer learning-based ensemble model for the classification of multiclass skin lesions, which provides satisfactory results. In the ensemble, two pre-trained models are combined to create a learning model that can capture rich representations of a variety of images. The proposed research could aid medical professionals in detecting seven different categories of skin ailments earlier. Traditionally, we used data augmentation to balance the dataset classes due to their imbalance. The HAM10000 dataset served as the basis for our experiments. The proposed IncepX-Ensemble model's performance was assessed based on four evaluation metrics in the study. Pre-processing of the images, the splitting of train and test sets, and the use of data enhancement techniques to balance the class data were all carried out prior to the training phase. The findings demonstrate that the improved ensemble transfer learning model outperformed the other suggested transfer learning-based methods, achiev-

ing an accuracy of 98%. For the HAM10000 data set, our suggested improved ensemble transfer learning model performs better in classification. Based on a comparison of the proposed model to a recent publication that was conducted on the same data set, IncepX-Ensemble has shown superior accuracy and a lower loss of information. In order to address the problem of data imbalance, we will use GAN-based synthetic data generation in future work to produce more realistic synthetic skin lesion images. Furthermore, we are going to develop a modified version of CycleGAN to generate skin lesion data using stain normalization methods.

REFERENCES

- [1] Marcos AM Almeida and Iury AX Santos. Classification models for skin tumor detection using texture analysis in medical images. *Journal of Imaging*, 6(6):51, 2020.
- [2] Michał H Strzelecki, Maria Strąkowska, Michał Kozłowski, Tomasz Urbaničzyk, Dorota Wielowieyska-Szybińska, and Marcin Kociotek. Skin lesion detection algorithms in whole body images. *Sensors*, 21(19):6639, 2021.
- [3] Gan Cai, Yu Zhu, Yue Wu, Xiaoben Jiang, Jiongyao Ye, and Dawei Yang. A multimodal transformer to fuse images and metadata for skin disease classification. *The Visual Computer*, 39(7):2781–2793, 2023.
- [4] Tanzila Saba, Muhammad Attique Khan, Amjad Rehman, and Souad Larabi Marie-Sainte. Region extraction and classification of skin cancer: A heterogeneous framework of deep cnn features fusion and reduction. *Journal of medical systems*, 43(9):289, 2019.
- [5] Giuseppe Argenziano, H Peter Soyer, Sergio Chimenti, Renato Talamini, Rosaria Corona, Francesco Sera, Michael Binder, Lorenzo Cerroni, Gaetano De Rosa, Gerardo Ferrara, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the internet. *Journal of the American Academy of Dermatology*, 48(5):679–693, 2003.
- [6] Saket S Chaturvedi, Jitendra V Tembhurne, and Tausif Diwan. A multi-class skin cancer classification using deep convolutional neural networks. *Multimedia Tools and Applications*, 79(39-40):28477–28498, 2020.
- [7] Ebrahim Mohammed Senan and Mukti E Jadhav. Analysis of dermoscopy

- 922 images by using abcd rule for early detection of skin cancer. Global 996
923 Transitions Proceedings, 2(1):1–7, 2021. 997
- [8] Yan Wang, Yangqin Feng, Lei Zhang, Joey Tianyi Zhou, Yong Liu, Rick 998
924 Siow Mong Goh, and Liangli Zhen. Adversarial multimodal fusion 999
925 with attention mechanism for skin lesion classification using clinical and 1000
926 dermoscopic images. Medical Image Analysis, 81:102535, 2022. 1001
- [9] Muhammad Attique Khan, Khan Muhammad, Muhammad Sharif, Tallha 1002
927 Akram, and Seifedine Kadry. Intelligent fusion-assisted skin lesion 1003
928 localization and classification for smart healthcare. Neural Computing and 1004
929 Applications, 36(1):37–52, 2024. 1005
- [10] Muhammad Nasir, Muhammad Attique Khan, Muhammad Sharif, 1006
930 Ikram Ullah Lali, Tanzila Saba, and Tassawar Iqbal. An improved strategy 1007
931 for skin lesion detection and classification using uniform segmentation 1008
932 and feature selection based approach. Microscopy research and technique, 1009
933 81(6):528–543, 2018. 1010
- [11] Mary Manandhar, Sarah Hawkes, Kent Buse, Elias Nosrati, and Veronica 1011
934 Magar. Gender, health and the 2030 agenda for sustainable development. 1012
935 Bulletin of the World Health Organization, 96(9):644, 2018. 1013
- [12] Recep Erol. Skin cancer malignancy classification with transfer learning 1014
940 University of Central Arkansas, 2018. 1015
- [13] Seena Joseph and Oludayo O Olugbara. Preprocessing effects on perfor- 1016
943 mance of skin lesion saliency segmentation. Diagnostics, 12(2):344, 2022. 1017
- [14] Rinaldi Pradhana Widhianto, Wahmisari Priharti, Fenty Alia, Afifah Risma 1018
944 Alfariyani, Zufar Asyraf Al-Hamid, and Eva Krishna Sutedja. Hardware 1019
945 design of skin cancer detection device. In 2023 International Conference 1020
946 on Engineering and Emerging Technologies (ICEET), pages 1–6. IEEE, 1021
947 2023. 1022
- [15] Huan Ding, Qirui Huang, and Ahmed Alkhayyat. A computer aided 1023
950 system for skin cancer detection based on developed version of the 1024
951 archimedes optimization algorithm. Biomedical Signal Processing and 1025
952 Control, 90:105870, 2024. 1026
- [16] Nauman Ullah Gilal, Samah Ahmed Mustapha Ahmed, Jens Schneider, 1027
953 Mowafa Househ, and Marco Agus. Mobile dermatoscopy: Class imbal- 1028
954 ance management based on blurring augmentation, iterative refining and 1029
955 cost-weighted recall loss. Journal of Image and Graphics, 11(2), 2023. 1030
- [17] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, 1031
956 Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew 1032
957 Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE 1033
958 conference on computer vision and pattern recognition, pages 1–9, 2015. 1034
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual 1035
962 learning for image recognition. In Proceedings of the IEEE conference on 1036
963 computer vision and pattern recognition, pages 770–778, 2016. 1037
- [19] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Wein- 1038
964 berger. Densely connected convolutional networks. In Proceedings of the 1039
965 IEEE conference on computer vision and pattern recognition, pages 4700– 1040
966 4708, 2017. 1041
- [20] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, 1041
968 Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, 1042
970 Mobilenets: Efficient convolutional neural networks for mobile vision 1043
971 applications. arXiv preprint arXiv:1704.04861, 2017. 1044
- [21] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for 1045
973 convolutional neural networks. In International conference on machine 1046
974 learning, pages 6105–6114. PMLR, 2019. 1047
- [22] Khalid M Hosny and Mohamed A Kassem. Refined residual deep 1048
975 convolutional network for skin lesion classification. Journal of Digital 1049
977 Imaging, 35(2):258–280, 2022. 1050
- [23] Khalid M Hosny, Wael Said, Mahmoud Elmezain, and Mohamed A 1051
979 Kassem. Explainable deep inherent learning for multi-classes skin lesion 1052
980 classification. Applied Soft Computing, 159:111624, 2024. 1053
- [24] Yousef S Alsahafi, Mohamed A Kassem, and Khalid M Hosny. Skin-net: a 1054
982 novel deep residual network for skin lesions classification using multilevel 1055
983 feature extraction and cross-channel correlation with detection of outlier. 1056
984 Journal of Big Data, 10(1):105, 2023. 1057
- [25] Kemal Polat and Kaan Onur Koc. Detection of skin diseases from 1058
986 dermoscopy image using the combination of convolutional neural network 1059
987 and one-versus-all. Journal of Artificial Intelligence and Systems, 2(1):80–1060
988 97, 2020. 1061
- [26] Daniel M Lima, Jose F Rodrigues-Jr, Bruno Brandoli, Lorraine Goeu- 1062
990 riot, and Sihem Amer-Yahia. Dermadl: advanced convolutional neural 1063
991 networks for computer-aided skin-lesion classification. SN Computer 1064
992 Science, 2:1–13, 2021. 1065
- [27] Md Kamrul Hasan, Md Toufick E Elahi, Md Ashraful Alam, Md Tasnim 1066
994 Jawad, and Robert Martí. Dermexpert: Skin lesion classification using 1067
995 hybrid convolutional neural network through segmentation, transfer learn- 1068
- ing, and augmentation. Informatics in Medicine Unlocked, 28:100819, 2022.
- [28] Peng Tang, Xintong Yan, Yang Nan, Shao Xiang, Sebastian Krammer, and Tobias Lasser. Fusionm4net: A multi-stage multi-modal learning algorithm for multi-label skin lesion classification. Medical Image Analysis, 76:102307, 2022.
- [29] Muhammad Attique Khan, Khan Muhammad, Muhammad Sharif, Tallha 1008
929 Akram, and Victor Hugo C de Albuquerque. Multi-class skin lesion 1009
930 detection and classification via teledermatology. IEEE journal of biomedical 1010
931 and health informatics, 25(12):4267–4275, 2021.
- [30] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data 1011
932 augmentation for deep learning. Journal of big data, 6(1):1–48, 2019.
- [31] Fábio Perez, Cristina Vasconcelos, Sandra Avila, and Eduardo Valle. Data 1012
933 augmentation for skin lesion analysis. In OR 2.0 Context-Aware Operating 1013
934 Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based 1014
935 Procedures, and Skin Image Analysis: First International Workshop, OR 1015
936 2.0 2018, 5th International Workshop, CARE 2018, 7th International 1016
937 Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held 1017
938 in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 1018
939 20, 2018, Proceedings 5, pages 303–311. Springer, 2018.
- [32] Zhiwei Qin, Zhao Liu, Ping Zhu, and Yongbo Xue. A gan-based image 1019
940 synthesis method for skin lesion classification. Computer Methods and 1020
941 Programs in Biomedicine, 195:105568, 2020.
- [33] Agnieszka Mikołajczyk and Michał Grochowski. Data augmentation 1021
942 for improving deep learning in image classification problem. In 2018 1022
943 international interdisciplinary PhD workshop (IIPhDW), pages 117–122. 1023
944 IEEE, 2018.
- [34] Debapriya Hazra, Yung-Cheol Byun, and Woo Jin Kim. Enhancing 1024
945 classification of cells procured from bone marrow aspirate smears using 1025
946 generative adversarial networks and sequential convolutional neural net- 1026
947 work. Computer Methods and Programs in Biomedicine, 224:107019, 2022.
- [35] SP Godlin Jasil and V Ulagamuthalvi. Deep learning architecture using 1027
948 transfer learning for classification of skin lesions. Journal of Ambient 1028
949 Intelligence and Humanized Computing, pages 1–8, 2021.
- [36] Adi Alhudhaif, Bandar Almaslukh, Ahmad O Aseeri, Osman Guler, and 1029
950 Kemal Polat. A novel nonlinear automated multi-class skin lesion 1030
951 detection system using soft-attention based convolutional neural networks. 1031
952 Chaos, Solitons & Fractals, 170:113409, 2023.
- [37] Saptarshi Chatterjee, Debangshu Dey, and Sugata Munshi. Integration 1032
953 of morphological preprocessing and fractal based feature extraction with 1033
954 recursive feature elimination for skin lesion types classification. Computer 1034
955 methods and programs in biomedicine, 178:201–218, 2019.
- [38] Khalid M Hosny, Mohamed A Kassem, and Mohamed M Fouad. Clas- 1035
956 sification of skin lesions into seven classes using transfer learning with 1036
957 alexnet. Journal of digital imaging, 33:1325–1334, 2020.
- [39] Nazia Hameed, Antesar M Shabut, Miltu K Ghosh, and M Alamgir 1037
958 Hossain. Multi-class multi-level classification algorithm for skin lesions 1038
959 classification using machine learning techniques. Expert Systems with 1039
960 Applications, 141:112961, 2020.
- [40] Zenghui Wei, Qiang Li, and Hong Song. Dual attention based network 1040
961 for skin lesion classification with auxiliary learning. Biomedical Signal 1041
962 Processing and Control, 74:103549, 2022.
- [41] Subhajit Chatterjee and Yung-Cheol Byun. Voting ensemble approach for 1042
963 enhancing alzheimer's disease classification. Sensors, 22(19):7661, 2022.
- [42] Ghadah Alwakid, Walaa Gouda, Mamoona Humayun, and Najm Us Sama. 1043
964 Melanoma detection using deep learning-based classifications. In Health- 1045
965 care, volume 10, page 2481. MDPI, 2022.
- [43] Anil Kumar Adepu, Subin Sahayam, Umarani Jayaraman, and Rashmika 1046
966 Arramraju. Melanoma classification from dermatoscopy images using 1047
967 knowledge distillation for highly imbalanced data. Computers in biology 1048
968 and medicine, 154:106571, 2023.
- [44] G Akilandasowmya, G Nirmaladevi, SU Suganthi, and A Aishwariya. 1049
969 Skin cancer diagnosis: Leveraging deep hidden features and ensemble 1050
970 classifiers for early detection and classification. Biomedical Signal 1051
971 Processing and Control, 88:105306, 2024.
- [45] Van-Dung Hoang, Xuan-Thuy Vo, and Kang-Hyun Jo. Categorical weight- 1052
972 ing domination for imbalanced classification with skin cancer in intelligent 1053
973 healthcare systems. IEEE Access, 2023.
- [46] Ghasem Shakourian Ghalejoogh, Hussain Montazery Kordy, and Farideh 1054
974 Ebrahimi. A hierarchical structure based on stacking approach for skin 1055
975 lesion classification. Expert Systems with Applications, 145:113127, 2020.

- [47] Zillur Rahman, Md Sabir Hossain, Md Rabiul Islam, Md Mynul Hasan,¹¹³⁰ and Rubaiyat Alim Hridhee. An approach for multiclass skin lesion classification based on ensemble learning. *Informatics in Medicine Unlocked*,¹¹³² 25:100659, 2021.
[48] Yi-Peng Liu, Ziming Wang, Zhanqing Li, Jing Li, Ting Li, Peng Chen, and¹¹³³ Ronghua Liang. Multiscale ensemble of convolutional neural networks for¹¹³⁴ skin lesion classification. *IET Image Processing*, 15(10):2309–2318, 2021.
[49] Cliff Rosendahl, Philipp Tschandl, Alan Cameron, and Harald Kittler.¹¹³⁵ Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic¹¹³⁶ pigmented lesions. *Journal of the American academy of dermatology*,¹¹³⁷ 64(6):1068–1073, 2011.
[50] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000¹¹³⁸ dataset, a large collection of multi-source dermatoscopic images of common¹¹³⁹ pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018.
[51] Subhajit Chatterjee, Debapriya Hazra, Yung-Cheol Byun, and Yong-Woon¹¹⁴⁰ Kim. Enhancement of image classification using transfer learning and gan-based¹¹⁴¹ synthetic data augmentation. *Mathematics*, 10(9):1541, 2022.
[52] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-¹¹⁴² Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE¹¹⁴³ conference on computer vision and pattern recognition, pages 248–255.¹¹⁴⁴ Ieee, 2009.
[53] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and¹¹⁴⁵ Zbigniew Wojna. Rethinking the inception architecture for computer¹¹⁴⁶ vision. In Proceedings of the IEEE conference on computer vision and¹¹⁴⁷ pattern recognition, pages 2818–2826, 2016.
[54] François Chollet. Xception: Deep learning with depthwise separable¹¹⁴⁸ convolutions. In Proceedings of the IEEE conference on computer vision and¹¹⁴⁹ pattern recognition, pages 1251–1258, 2017.
[55] Muhammad Attique Khan, Tallha Akram, Muhammad Sharif, Seifedine¹¹⁵⁰ Kadry, and Yunyoung Nam. Computer decision support system for skin¹¹⁵¹ cancer localization and classification. *Computers, Materials & Continua*,¹¹⁵² 68(1), 2021.
[56] Amin Tajerian, Mohsen Kazemian, Mohammad Tajerian, and Ava Akhavan¹¹⁵³ Malayeri. Design and validation of a new machine-learning-based¹¹⁵⁴ diagnostic tool for the differentiation of dermatoscopic skin cancer images.¹¹⁵⁵ *Plos one*, 18(4):e0284437, 2023.
[57] Parvathaneni Naga Srinivasu, Jalluri Gnana SivaSai, Muhammad Fazal¹¹⁵⁶ Ijaz, Akash Kumar Bhoi, Wonjoon Kim, and James Jin Kang. Classification¹¹⁵⁷ of skin disease using deep learning neural networks with mobilenet¹¹⁵⁸ v2 and lstm. *Sensors*, 21(8):2852, 2021.
[58] Karl Thurnhofer-Hemsi and Enrique Domínguez. A convolutional neural¹¹⁵⁹ network framework for accurate skin cancer detection. *Neural Processing*,¹¹⁶⁰ Letters, 53(5):3073–3093, 2021.
[59] Ioannis Kousis, Isidoros Perikos, Ioannis Hatzilygeroudis, and Maria¹¹⁶¹ Virvou. Deep learning methods for accurate skin cancer recognition and¹¹⁶² mobile application. *Electronics*, 11(9):1294, 2022.
[60] Emrah Çevik and Kenan Zengin. Classification of skin lesions in dermatoscopic¹¹⁶³ images with deep convolution network. *Avrupa Bilim ve Teknoloji*,¹¹⁶⁴ Dergisi, pages 309–318, 2019.
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179



computing.

DR. JOON-MIN GIL received a Ph.D. degree in Computer Science and Engineering from Korea University, Korea in 2000. He is currently a professor at the Department of Computer Engineering, Jeju National University (JNU), Jeju, Korea. Before joining JNU in Sept. 2023, he was with Daegu Catholic University as a professor from Mar. 2006 to Aug. 2023. His recent research interests include cloud computing, big data computing, artificial intelligence, and distributed



DR. YUNG-CHEOL BYUN received his Ph.D. and MS from Yonsei University in 1995 and 2001 respectively, and BS from Jeju National University in 1993. He worked as a special lecturer in SAMSUNG Electronics and SDS from 1998 to 2001. From 2001 to 2003, he was a senior researcher at the Electronics and Telecommunications Research Institute (ETRI). He was promoted to join Jeju National University as an assistant professor in 2003. Now, he is a full professor in the Computer Engineering Department at the University. His research interests include the areas of AI & machine learning, pattern recognition, blockchain and deep learning-based applications, big data and knowledge discovery, time-series data analysis and prediction, image processing, medical image applications, and recommendation systems.

• • •



SUBHAJIT CHATTERJEE has done bachelor's in computer application (BCA) in 2012. He completed his master's in computer application (MCA) in the year 2015. He is currently pursuing the Ph.D. degree with the Machine Learning Laboratory, Department of Computer Engineering, Jeju National University, South Korea. He Worked related to information technology, software engineering and customer relationship management of the banking sector for five years. His research interests include artificial intelligence, deep learning, convolutional neural network, image processing.