Skin Cancer Diagnosis under Intra-Class Variability Using Transfer Learning and Fuzzy Logic-Based Decision Refinement

Zahraa Saad Abdul Wahid, Zahraa Ch. Oleiwi and Rasha Falah Kadhem
College of Computer Science and Information Technology, University of Al-Qadisiyah, 58001 Al-Diwaniyah, Iraq
zahraasaadadbulwahid@gmail.com, zahraa.chaffat@qu.edu.iq, rasha.kadhem@qu.edu.iq

Keywords: Skin Cancer, Transfer Learning, Intra-Class Variability, CNN, Pre-Trained Networks, Fuzzy Logic.

Abstract: Skin cancer is a disease that is most prevalent and life threatening globally, where early diagnosis is very important in enhancing the survival chances of the patient. Nevertheless, the automation of dermoscopic image classification is still problematic because of the high similarity of images of various lesion types and the large inter-class variation. The Structural Similarity Index Measure (SSIM) was used to measure this difficulty on the 3297 images of the Dermoscopic skin lesion dataset, and the results showed that the similarity score (SSIM) was low among several samples in the same category, indicating the difficulty of the classification task. To overcome such problems, a hybrid deep-learning architecture is proposed to classify skin cancer, which uses various convolutional neural networks and fuzzy logic reasoning. Three pre-trained models, namely, DenseNet201, EfficientNetV2B2, and ResNet50, were used to extract features with the help of transfer learning. A number of hybrid models were developed by fusing the feature vectors of these networks. The experimental findings indicate that the tri-network hybrid model, which combines all three models, is better than each of the models and their combination. The proposed model had the overall accuracy of 0.91 and class-wise recall of 0.88 and 0.95, as well as the F1-score of 0.91 and 0.90 in benign and malignant classes, respectively. Moreover, the integration of a fuzzy logic inference system to narrow the prediction confidence enhanced the strength of the ultimate decision.

1 INTRODUCTION

One of the most common and life-threatening illnesses on earth is skin cancer, whose rising number is due to environmental, genetic and lifestyle factors [1]. Early and precise diagnosis is important in enhancing survival rates and low treatment costs. Nonetheless, other traditional diagnostic processes, including visual inspection and biopsy, are usually time-consuming, subjective and extremely reliant on the skills of dermatologists [1]. Thus, artificial intelligence computer-aided diagnostic systems have become the hopeful instruments of the early detection [2], [3].

Recent developments in the field of deep learning, especially convolutional neural networks (CNNs), have led to a dramatic increase in the ability of automated skin lesion classification systems. These models can capture more complicated visual characteristics such as color, texture and lesion boundaries, so that they can better distinguish benign and malignant lesions. Additionally, other researchers have also shown that deep learning models, which are trained on large collections of data, including ISIC and HAM10000, can be more effective than classical machine learning methods [4]-[6].

To further improve performance, more recent studies have been done on hybrid and ensemble models, which combine multiple deep learning architectures or on deep learning combined with other intelligent methods [7], [8]. These are methods to enhance feature representation, minimize overfitting, and optimize generalization [9]-[11].

Moreover, novel architectures, such as EfficientNet, ResNet, and transformer-based models, have been extensively studied, and methods to handle the challenges of limited data, class imbalance, and transfer learning, data augmentation, and attention have been explored [12].

Although such progress has been made, there are still a number of challenges. Most of the currently available methods use single CNN models, which might not be effective in capturing the complicated patterns of dermoscopic images, especially when there is high intra-class variability [13]. More so, dermoscopic datasets are mostly highly variable within the same category, where visually dissimilar images can be of the same diagnostic category. This difficulty is measurably illustrated in this paper with the Structural Similarity Index Measure (SSIM), wherein the similarity scores (SSIM less than 0.2) between large numbers of samples of identical type is extremely low. Also, the majority of the models only use probabilistic outputs, which do not necessarily reflect the uncertainty in borderline cases.

To overcome these shortcomings, unlike traditional methods that use one CNN or standard ensemble strategies, the proposed model presents a tri-network hybrid model that teaches feature-level fusion of DenseNet201, EfficientNet and ResNet50. The obtained feature vectors are then concatenated and input into a single classification head, which allows the model to learn more discriminative representations. The primary contributions of this work may be outlined as follows.

1) Quantitative analysis of intra-class variability using SSIM, demonstrating the complexity of dermoscopic image classification.
2) Development of a multi-network hybrid architecture based on DenseNet201, EfficientNetV2B2, and ResNet50 for enhanced feature representation.
3) Construction and evaluation of multiple hybrid models, including pairwise and tri-network combinations.
4) Proposal of a tri-network hybrid model that achieves superior performance compared to individual and pairwise models.
5) Integration of a fuzzy logic system to improve decision reliability and handle uncertainty in classification.

Comprehensive experimental evaluation demonstrating the effectiveness of the proposed approach, achieving an accuracy of approximately 0.91.

2 RELATED WORKS

Some of the studies have used transfer learning, hybrid CNN models and ensemble techniques to enhance the classification accuracy where there is sparse medical data. Regardless of the encouraging reports in the literature, the reliability of automated diagnostic systems is still hampered by a number of challenges. A large intra-class variability of dermoscopic lesions is one of the most serious challenges because lesions of the same type may have large differences in appearance in terms of color, texture and shape. Thus, scholars have tried to optimize the classification performance by using better feature extraction, data augmentation, and hybrid deep learning models [14]. The present section summarizes some of the recent research concerning the classification of skin cancer by using dermoscopic images and identifies their strengths and weaknesses.

The proposed AI-based framework was a deep learning model that utilized an image transfer learning model called DenseNet-121 to detect skin cancer (Hamsalekha et al., 2024). A custom dermoscopic dataset comprising 2000 images was used to train the model, with 1500 images being used as the training sample, and 500 images serving as the testing sample. To enhance generalization, a number of preprocessing methods were incorporated like image resizing, normalization and data augmentation approach like horizontal flipping and random cropping. The DenseNet-121 model was trained on ImageNet, and the original classification layer was adjusted to binary classification between benign and malignant lesions. The test outcomes indicated that the developed model had 91% accuracy, 0.89 precision, 0.91 recall, and 0.90 F1-score [15].

Hussein et al. (2025) suggested a hybrid deep learning model to classify skin cancer images based on using a Hybrid Quantum Convolutional Neural Network (HQCNN) with the BiLSTM and MobileNetV2 frameworks. The objective of the study was to deal with the complexity of the skin lesion classification due to minor differences in the visual appearance between benign and malignant lesions. In the suggested method, HQCNN model was used to extract a hierarchical spatial representation of dermoscopic images, and MobileNetV2 was adopted to extract features efficiently. Also, BiLSTM networks were included to calculate relationships to the context and extract sequential dependencies in the obtained features, which facilitated better learning of representations. Experiments were performed on a clinically relevant image database of skin cancer to test the proposed framework between two resolution images (32x32 and 128x128 pixels) to determine how image size affects classification performance. The results of the experiment revealed that the hybrid HQCNN-BiLSTM-MobileNetV2 model had an accuracy of 0.977 and 0.893 in training and testing, respectively on 128x128 color images. Moreover, the model had an F1-score of 0.8981 and a recall of 0.9433 of malignant cases, which means that the model is highly sensitive to cancer and it has good diagnostic accuracy [16].

Asgharnezhad et al. (2025) have performed an in-depth investigation of automated skin cancer classification based on deep learning models with uncertainty quantification (UQ) methods. The paper was intended to enhance reliability of automated diagnostic systems, which are inherently subject to the shortcomings of the traditional deep learning models, namely, the impossibility of quantifying uncertainty of prediction. The experiment was performed the HAM10000 dataset, and the dataset consists of dermoscopic images of different categories of lesions on the skin. During the initial research, various pre-trained image feature extractors were tested in a transfer learning system, among which are CLIP-based vision transformers, ResNet50, DenseNet121, VGG16, and EfficientNet-V2 Large. These extractors of the features were integrated with the conventional machine learning classifier like the Support Volume Machine (SVM), extreme Gradient Boosting (XGBoost) and the Logistic Regression to evaluate their capacity to classify. The findings showed that CLIP-based vision transformer models, especially LAION CLIP ViT-H/14 together with SVM performed optimally in the classification task compared to the tested methods. At the second step, the authors added uncertainty-aware methods of learning, such as Monte Carlo Dropout (MCD), ensemble learning, and Ensemble Monte Carlo Dropout (EMCD) to determine the accuracy of model predictions. In order to measure uncertainty estimation effectiveness, the research proposed a number of uncertainty-sensitive measures of evaluation, such as uncertainty accuracy (UAcc), uncertainty sensitivity (USen), uncertainty specificity (USpe), and uncertainty precision (UPre). The outcomes of the experiment proved that the balance between the classification accuracy and the estimation of uncertainty have a trade-off provided by the ensemble-based approaches, whereas the EMCD method is more sensitive to uncertain predictions [17].

Gouda et al. (2022) introduced a deep-learning-based method of automated detection of skin cancer in dermoscopic lesions of the skin using the images of skin lesions. The research problem was to differentiate the two main types of tumors benign and malignant with the help of convolutional neural network models. The experiments were performed with the use of the ISIC 2018 dataset consisting of 3533 images of skin lesions of various types. At the preprocessing phase, the authors used Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) to boost the quality of images. Other preprocessing operations were performed as well such as data augmentation, normalization, and resizing of the images to enhance the generalization of the models and the effective size of the dataset. To classify it, the research tested a number of deep learning structures. Besides a hand-crafted CNN model, several transfer learning networks were also used, i.e. ResNet50, InceptionV3, and Inception-ResNet. The ISIC 2018 dataset was used to fine-tune these models to enhance their classification. By experimentation, InceptionV3 demonstrated the highest performance of 0.858 and the next best performance of 0.84 by Inception-ResNet and 0.837 by ResNet50 and the proposed CNN model. The authors emphasized that the addition of ESRGAN at the preprocessing step helped in enhancing the quality and classification of images[18].

Hosseinzadeh et al. (2024) recommended hybrid architecture to classify skin cancer, which was developed based on the deep learning feature extraction and the ensemble machine learning methods. The main goal of the research was to enhance the accuracy of the skin lesion classification method through the combination of varying feature extraction and feature selection algorithms. Some transfer learning models, including DenseNet-201, were utilized in their approach where it was selected as the primary feature extractor of the backbone. The discriminative power of the extracted features was enhanced by using various techniques of feature selection, among them being Univariate selection, Mutual Information, ANOVA, PCA, XGBoost, Lasso, Random Forest and Variance-based selection. Once the features had been selected the reduced feature sets were inputted into various machine learning classifiers including Multi-Layer Perceptron (MLP), extreme Gradient Boosting (XGBoost), Random Forest (RF), and Naive Bayes (NB). Lastly, the ensemble learning strategies were used to integrate the predictions of the most effective classifiers. The accuracy and sensitivity were reported as 87.72% and 92.15% respectively, and this showed that addition of deep feature extraction to ensemble learning were effective in skin cancer classification [19].

The feature extraction and optimization methods proposed by Priya and Rajeswari (2024) in their machine learning-based framework are handcrafted methods used to identify skin cancer. The purpose of the study was to enhance the discrimination of the skin lesion patterns with sophisticated texture and shape features analysis. Firstly, the authors used Contourlet Transform (CT) and Local Binary Pattern (LBP) techniques to extract structural and texture features of dermoscopy images, that is, lesion edges, contrast differences, and morphological features. Particle Swarm Optimization (PSO) was employed to choose the most informative features and minimize computation time as the extracted features created the high-dimensional feature space. A number of machine learning classifiers, such as Support Vector Machine (SVM), Random Forest (RF) and Neural Networks (NN), were then trained using the optimized feature set. The experimental findings showed that the proposed solution had a precision of 86.9, and SVM and Neural Network classifiers had shorter training time than the Random Forest [20].

Hasan et al. (2019) suggested a deep learning system which detects skin cancer using dermoscopic images in an automated fashion. The analysis aimed at enhancing the early diagnosis in the form of an artificial intelligence-based classification framework. The given pipeline initially utilized the image segmentation and feature extraction methods to determine the useful features in the affected areas of the skin. A Convolutional Neural Network (CNN) classifier was then used to do the final task of skin lesion classification. It was experimentally proven that the proposed CNN-based system reached the classification accuracy of 89.5% and the training accuracy of 0.937 when tested on a publicly available skin lesion data set. The authors emphasized that automated diagnosis of skin cancer with the usage of deep learning models could be effectively supported by the hierarchy of features learning [21].

Meghatria et al. (2026) explored curriculum learning as an approach in a framework of deep learning melanoma detection. The study aim was to ensure the strength of skin lesion classification models through the gradual improvement of the training process using the simplest to the most complex samples. A number of transfer learning architectures, such as VGG16, ResNet50, and EfficientNetB0, were evaluated using the proposed framework and trained on dermoscopic images of the ISIC 2019 and 2020 dataset. Bagging and other ensemble learning methods were employed to aggregate several model predictions so as to enhance classification performance. In addition, the experiment included curriculum learning, in which the samples of training were presented with the models getting harder and harder to enhance the generalization of the model. The experimental performance demonstrated that the curriculum learning strategy used in EfficientNetB0 was the most successful in terms of the highest F1-score (0.9077) than the traditional methods of fine-tuning and ensemble approaches [22].

Despite the above-mentioned studies showing that there are great advances in automated classification of skin cancer with machine learning, deep learning, and ensemble approach, there are still a few limitations. Majority of the current methods utilize either single deep learning models or standard ensemble methods, and do not attempt to fuse features across multi- CNN architectures to obtain complementary image representations. Moreover, the ultimate classification decision made in such approaches is normally based on the model outputs directly without integrating any intelligent decision refinements like fuzzy logic which can enhance predictability in any doubtful situation. Besides, the problem of high intra-class variability in dermoscopic images, where both benign and malignant lesions can display extremely similar appearances, is not explicitly discussed in a number of studies. Such shortcomings underscore the necessity of stronger frameworks, which combine multi-network feature fusion, hybrid deep learning designs, and fuzzy-based decision enhancement to gain classification stability and diagnostic accuracy. To overcome these shortcomings, the current study suggests a hybrid CNN model by combining several deep neural networks via feature fusion, and a fuzzy logic used to refine the decision-making process to enhance the resilience of the classification in case of intra-class variability of dermoscopic skin lesion images. A summary of recent studies and their key limitations is presented in Table 1.

3 PROPOSED METHOD

The presented structure enhances the classification of skin cancer by integrating various deep learning models with decision refinement, based on fuzzy logic. The general workflow consists of multiple steps, such as the preparation of data, preprocessing, deep feature extraction with different CNN models, hybrid feature fusion, model training and evaluation, and, lastly, the support of decision using the fuzzy logic. Figure 1 shows the entire structure of the proposed system.

3.1 Dataset Preparation and Label Encoding

The dermoscopic image paths and the corresponding class labels were read and loaded into the dataset. As the class labels were originally in the form of textual categories denoting benign and malignant lesions, a label encoding process was used to turn them into numerical values that could be used by the deep learning models. Particularly, benign lesions had been coded as 0, and malignant lesions had been coded as 1.

Table 1: Summary of recent studies on skin cancer classification.

Study	Methodology	Key Limitations
Hamsalekha et al. (2024) [15]	DenseNet-121 for binary skin lesion classification (benign vs. malignant)	Small size of datasets (2000 images) which could be a limitation to generalization. Besides, the research fails to present the class-wise error analysis in detail by classes as a way of showing model robustness.
Hussein et al. (2025) [16]	Hybrid model combining HQCNN, BiLSTM, and MobileNetV2	Complex models that are characterized by relatively low performance. The hybrid design lacks the ability to identify discriminative features to medical images and preprocessing procedures can compromise on the quality of the image.
Asgharnezhad et al. (2025) [17]	Multiple pre-trained feature extractors (CLIP, ResNet50, DenseNet121, VGG16, EfficientNet-V2) with ML classifiers (SVM, XGBoost, Logistic Regression)	Adding many feature extractors does not make the use of these extractors proportionately more effective. Deep visual patterns may not be completely represented in traditional machine learning classifier.
Gouda et al. (2022) [18]	CNN-based classification with multiple transfer learning models (ResNet50, InceptionV3, Inception-ResNet)	Repeated training and the use of a number of models leads to high computational complexity. Minimal attention to effective feature representation to enhance the classification performance.
Hosseinzadeh et al. (2024) [19]	DenseNet201 with Lasso feature selection and ML-based ensemble methods	Absence of a deep feature fusion. The extensive use of traditional machine learning ensembles with no full utilization of hybrid deep representations.
Priya & Rajeswari (2024) [20]	Handcrafted features (CT, LBP) with PSO optimization and ML classifiers	Handcrafted features may not adequately capture intra-class variability in dermoscopic images, despite dimensionality reduction.
Hasan et al. (2019) [21]	CNN-based feature extraction and classification	Poor capability of feature extraction and inadequate generalization performance because of absence of advanced hybrid strategies or fusion strategies.
Meghatria et al. (2025) [22]	Curriculum learning with ensemble deep learning	Emphasis should be on feature-level fusion instead of ensemble methods that can restrict the capability to deal with variability of skin lesion images.

3.2 Data Preprocessing

A preprocessing operation was utilized to have a consistent input. All dermoscopic data were down sampled to 224x224 pixels, which is the size of the input demanded by the CNN structures used. Pixel values were scaled to enhance the stability of training. Also, data augmentation methods such as flipping were used to enrich the data and eliminate overfitting.

3.3 Dataset Variability Analysis

The analysis of the dataset variability was performed prior to the training of the classification models to gain a better idea concerning the problems related to the classification of the skin lesions. Dermoscopic images can be characterized by large intra-class variability, as samples of one diagnostic group could have significant differences in terms of color distribution, patterns of the texture, and delimiting lesions. In order to measure this variability, the Structural similarity Index (SSIM) was applied to estimate the similarity of samples of the same category via (1):

$$SSIM(X,\mathrm{Y})=\frac{(2\mu_X\mu_Y+C_1)(2\sigma_{XY}+C_2)}{(\mu_X^2+\mu_Y^2+C_1)(\sigma_X^2+\sigma_Y^2+C_2)}. \quad (1)$$

The constants C_1 and C_2 are defined as follows:

$$C_1=(K_1L)^2,\ C_2=(K_2L)^2.$$

K_1 and K_2 are significant constants, while L equals 255, the maximum pixel value [23].

The comparison indicated that most of the samples had significantly low similarity values, as they often fell below 0.2 which implied high visual diversity in the same class. This variability may adversely influence the performance of deep learning classifiers, with one network architecture being unable to learn all the discriminative patterns in the dataset. This observation inspired the adoption of hybrid CNN architectures, i.e., multi networks are fused together with the aim of learning complementary feature representations that can be more robust to this variability.

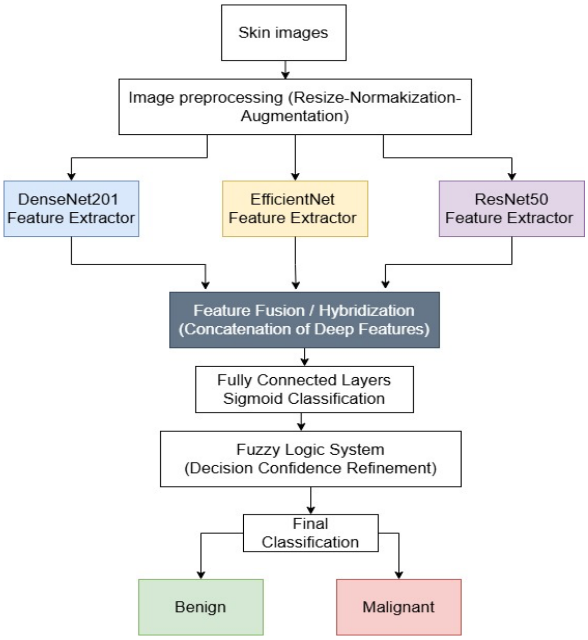

Figure 1: The complete architecture of the proposed system.

3.4 Deep Learning Models

The three prepared convolutional neural networks that are used in the proposed framework are DenseNet201, EfficientNetV2B2, and ResNet50 [24], [25]. The use of transfer learning was taken to utilize the information gathered in the large scale datasets[26], [27]. The feature extractors that were used were the convolutional layers of the pre-trained models and the final classification layers were adjusted to the binary skin lesion classification task. In particular, the last fully connected layers of both networks were eliminated by setting the parameter include top to False so that the models could be used as deep feature extractors of dermoscopic images whose size was 224 x 224 x 3.

The architectures that were selected were selected due to their complementary capabilities in medical image analysis. DenseNet201 is distinguished by high connectivity, thus the feature reuse, and the mitigation of the vanishing gradient issue, which is especially useful in small datasets. ResNet50 has good residual learning properties that enable the training of deeper representations. EfficientNetV2B2 has a scalable architecture with accuracy and computation efficiency. DenseNet201 has been shown to be very generalized and accurate when the amount of data is low in medical imaging tasks but EfficientNetV2B2 can be more efficient with adequate data and the appropriate scale. Hence, a combination of these architectures enables the model to take advantage of the different and complementary representations of features.

The weights of about 40 percent of the initial layers of the networks were frozen, and the deeper layers were trained. The decision to freeze 40% of the layers was a compromise between the desire to retain pre-trained features and the need to perform adequate model adaptation. The deep feature maps were then extracted and then subjected to a Global Average Pooling layer, which subsidiates the spatial dimensions of the feature maps and the most important global information is retained. This was followed by a sequence of fully connected layers to learn more deep features of the extracted features. Namely, a 64-nerve and 128-nerve dense layer was employed with the LeakyReLU activation function in order to add non-linearity and enhance the adaptability to learn features. The dropout layers with the rate of 0.2 were added between the dense layers to decrease the chances of overfitting. Lastly, the output layer was implemented using a sigmoid activation function to classify between benign and malignant skin lesions and was applied in binary classification. General transfer learning and fine-tuning procedure used in the utilized CNN models is summarized in Algorithm 1.

Input: Dermoscopic image dataset.
Output: Fine-tuned CNN model.

1. Load pre-trained CNN architectures;
2. Remove the original classification layers (include_top ← False);
3. Set input image size to 224 × 224 × 3;
4. Compute number of layers to freeze:
   $n \leftarrow 0.4 \times \text{total\_layers}$
5. Freeze the first n layers of each network;
6. Set remaining layers as trainable;
7. Apply GlobalAveragePooling2D;
8. Apply Dense layer (64 neurons, LeakyReLU activation);
9. Apply Dropout layer (rate ← 0.2);
10. Apply Dense layer (128 neurons, LeakyReLU activation);
11. Apply Dropout layer (rate ← 0.2);
12. Apply Output layer (Sigmoid activation);
13. Train the model;
14. Evaluate the model.

Algorithm 1: Transfer learning and fine-tuning procedure.

3.5 Hybrid Model Construction

A hybrid models were built by pooling deep features of various CNN architectures, so the ability of the features representation can be enhanced. The networks were then sending combining feature vectors through to the final classification layers.

There were two hybrid models explored:

- Pairs of hybrid models with the characteristics of two networks.
- A tri-network hybrid architecture that has the characteristics of DenseNet201, EfficientNetV2B2, and ResNet50 all at the same time.

The fourth hybrid model, which combines all three architectures, was proved to be superior in performance over other hybrid models.

The hybrid model is based on the assumption that the complementary feature-extracting power of various CNN models can be harnessed. Two CNN models trained in advance are used in this method as parallel feature extractors. The processes of each network take the same input dermoscopic image and produce a high-level feature map. These feature maps are then run through a layer of Global Average Pooling to transform spatial feature maps into compact feature vectors. Once the features have been extracted the feature vectors of the two, three or more networks are concatenated to create a single feature vector which is a combination of the features learned by the multiple networks. This aspect fusion approach allows the model to be more discriminative in contrast to a network. The combination feature is then put through a fully connected neural network with a series of dense layers using LeakyReLU activation functions. To alleviate overfitting and enhance the model generalization, dropout layers are used in between the dense layers. Lastly, the output layer has a sigmoid activation function to generate the likelihood of the input image being part of the malignant category. This hybrid design enables the model to enjoy the complementary advantages of several CNN feature extractors; hence, leading to enhanced classification outputs. The diverse steps of the process of hybrid feature fusion are encapsulated in Algorithm 2.

3.6 Model Training, Evaluation and Testing

Nadam optimizer was utilized to train the models on Binary cross-entropy loss. The constant batch size of 32 and learning rate of 0.001 was set. The epochs were determined to be 6 by experimental results, with the model converging early and further epochs (12 and 25) not providing any significant gain at a cost of increased numbers of epochs. The efficiency of the suggested models was measured with a number of standard classification tools, such as accuracy, precision, recall, F1-score, and confusion matrix analysis. These measures are a complete evaluation of the model in terms of the correct classification of benign and malignant lesions.

Input: Dermoscopic image x.
Output: Classification probability.

1. Load pre-trained N CNN models;
2. For each N^{th} CNN models do:
3. Extract feature maps;
4. feature vector V_{Nt} ← Apply Global Average Pooling(Extracted feature maps);
5. End for;
6. Concatenate feature vectors (V_1, V_2,….V_N);
7. Add Dense layer (Units←512, Activation function←LeakyReLU);
8. Apply Dropout (0.5);
9. Add Dense layer (Units←256, Activation function←LeakyReLU);
10. Apply Dropout (0.3);
11. Add Sigmoid output layer;
12. Compute classification probability.

Algorithm 2: Hybrid CNN feature fusion.

3.7 Fuzzy Logic Decision Enhancement

A Fuzzy Logic decision enhancement module was also added to the proposed framework to improve the reliability of the classification decision. Although the hybrid CNN models generate a probability score that expresses the probability that the class can be malignant, there can still be uncertainty in the outputs of the probabilistic models, especially in the borderline cases. Thus, fuzzy inference system was implemented to narrow down on the final decision. The cnn_pro input variable is a variable which has a range of [0,1] and is modeled with three membership functions: low and high which are trapezoidal functions and medium which is a triangular function. In the same way, trapezoidal functions are used to model the output variable (malignancy) of the two classes of benign and malignant and a triangular function of the suspicious category. Three fuzzy rules are utilized:

- In case the probability is low, then the output is benign.
- In case of a medium probability, then the output is suspicious.
- In case the probability is high, then the output is malignant.

Defuzzification through the centroid method is applied to the fuzzy inference process to give a continuous score of malignancy. Lastly, the output is converted into a binary by a suitable tuned threshold.

The fuzzy system used to minimize false negatives, which are essential in medical diagnosis as they can cause. The fuzzy module is, therefore, improves sensitivity toward diseased cases, making it a clinically meaningful enhancement rather than a purely numerical improvement. The fuzzy decision enhancement process includes step-by-step processes, which are outlined in Algorithm 3.

```
Input: CNN probability predictions P.
Output: Final binary classification.
Define fuzzy input and output variables;
Define input membership functions (low, medium, high);
Define output membership functions (benign, suspicious, malignant);
Construct fuzzy rule base;
If (cnn_prob is low) Then;
    Output ← benign;
Else If (cnn_prob is medium) Then;
    Output ←suspicious;
Else If (cnn_prob is high) Then;
    Output ←malignant;
End If;
Initialize fuzzy inference system;
For each probability p in P do:
p ←fuzzy input variable;
    Apply fuzzy inference;
    S ←Compute defuzzified score;
    Assign class label based on threshold;
      If (S≥ threshold) Then;
          class ← 1 (malignant);
      Else;
          class ← 0 (benign);
      End If;
End for.
```

Algorithm 3: Fuzzy logic decision enhancement.

4 EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

This section presents the experimental findings that will be used to assess the efficacy of the proposed hybrid deep learning model in the classification of skin cancer. In order to have a comprehensive analysis, seven models were modeled and reviewed. They are three single CNN architecture, three hybrid architecture, based on the combination of two networks, and a last hybrid architecture that combines the characteristics of all three networks. The experimental analysis is thus made to consist of a comparison that is made between the results of the experiment prior and subsequent to the imposition of the fuzzy logic, with the aim of investigating its effect on the performance of classification.

4.1 Experimental Setup

The experiments were done based on the publicly available Skin Cancer: Malignant vs. Benign dataset on the Kaggle. The information is a dataset consisting of dermoscopic images of skin lesions with two classes benign and malignant. The images were first obtained in the ISIC Archive and then down sampled to 224 x 224 pixels, which is the input size of the convolutional neural networks in this paper.

The dataset is divided into two major subsets, a training set, on which the model should be learned, and a testing set, on which the performance of a model should be tested. Training and testing sets were divided into 80:20. Overall, there are 2637 training and 660 testing images with benign and malignant lesions, respectively. The class distribution is approximately balanced, where the training set and the testing set have roughly equal proportions of benign and malignant samples (54.6 and 45.4% respectively). The distribution of the dataset employed in the experiments is as follows as shown in Table 2. Based on this dataset, a number of CNN-based models and hybrid ones were trained and tested to examine the effects of feature fusion and enhancement of fuzzy decisions on the performance of skin cancer classification.

4.2 Quantitative Evaluation

Three separate individual convolutional neural network models were trained on the skin cancer dataset to determine the accuracy of the models when used as the baseline: DenseNet201, ResNet50 and EfficientNetV2B2. The individual CNN architectures had varying performance as indicated by the DenseNet201 model that had the highest cumulative success as indicated in Table 3. Particularly, the accuracy rate of DenseNet201 is 0.8985 exorcistically above the results of both ResNet50 and EfficientNetV2B2 that obtained 0.7985.

On the benign class, DenseNet201 obtained a precision of 0.8786 and a recall of 0.9444. Likewise, with the malignant group, the model had precision of 0.9267 and a recall of 0.8433 leading to F1-score of 0.8831. On the contrary, both ResNet50 and EfficientNetV2B2 showed lower performance especially on the detection of malignant lesions where the recall value had decreased to 0.6733. This indicates a substantial trend of malignant lesions that were misclassified compared to DenseNet201. The dense connections in DenseNet201 enable strong feature reuse. This facility enhanced accuracy, particularly with medical images as compared to ResNet50 and EfficientNetV2B2.

The final decision is optimized by the fuzzy inference system. The findings of Table 4 indicate that the use of fuzzy logic marginally enhanced the accuracy of the classification of the analyzed models.

DenseNet201 model has an accuracy of 0.90, which is slightly higher than the results of the baseline when fuzzy logic is not used. On the same note, ResNet50 and EfficientNetV2B2 recorded significant gains in terms of their classification abilities with both of the models having an accuracy of 0.82 upon fuzzy logic application. Specifically, the recall of the malignant class rose to 0.87 compared to 0.6733, which means that the fuzzy inference system at least partially contained the number of malignant cases that were left out. This is based on the fact that fuzzy logic is capable of dealing with the uncertainty in CNN probability output, particularly when dealing with borderline samples whose model certainty is therefore not clear. Although the overall accuracy improvement is not very high, the fuzzy decision layer drastically improves the detection of malignant lesions to a large extent where false negatives are minimized and recall values are increased, which is of great essence in medical diagnosis uses. Despite the fact that DenseNet201 performed the best among the individual CNN models, the performance still shows that there are some weaknesses in the way variability of dermoscopic image is captured.

Table 2: Dataset distribution.

Dataset Split	Benign	Malignant	Total
Training Set	1440	1197	2637
Testing Set	360	300	660
Total	1800	1497	3297

Table 3: Performance of individual CNN models for skin lesion classification without fuzzy logic enhancement.

Class	Evaluation measure	Dense Net201	Res Net50	Efficient NetV2B2
Benign(0)	Precision	0.8786	0.7683	0.7683
	Recall	0.9444	0.9028	0.9028
	F1-score	0.9103	0.8301	0.8301
Malignant(1)	Precision	0.9267	0.8523	0.8523
	Recall	0.8433	0.6733	0.6733
	F1-score	0.8831	0.7523	0.7523
Macro average	Precision	0.9026	0.8103	0.8103
	Recall	0.8939	0.7881	0.7881
	F1-score	0.8967	0.7912	0.7912
Weighted average	Precision	0.9005	0.8065	0.8065
	Recall	0.8985	0.7985	0.7985
	F1-score	0.8979	0.7948	0.7948
Accuracy		0.8985	0.7985	0.7985

Table 4: Performance of individual CNN models after applying fuzzy logic decision enhancement.

Class	Evaluation measure	Dense Net201	Res Net50	Efficient NetV2B2
Benign(0)	Precision	0.89	0.88	0.88
	Recall	0.93	0.77	0.77
	F1-score	0.91	0.82	0.82
Malignant(1)	Precision	0.92	0.76	0.76
	Recall	0.86	0.87	0.87
	F1-score	0.89	0.81	0.81
Macro average	Precision	0.90	0.82	0.82
	Recall	0.90	0.82	0.82
	F1-score	0.90	0.82	0.82
Weighted average	Precision	0.90	0.82	0.82
	Recall	0.90	0.82	0.82
	F1-score	0.90	0.82	0.82
Accuracy		0.90	0.82	0.82

This fact is why some of their models, including ResNet50 and EfficientNetV2B2, had reverse performances, with the greatest difficulty in detecting malignant cases, which are indicated by the comparatively low recall rates prior to the use of fuzzy logic. Although, the fuzzy decision module enhanced the robustness of the final decision, the findings indicate that not all dermoscopic images might be represented by only one CNN architecture. To resolve this drawback, a hybrid approach has been adopted, where various CNN architectures acquire complementary feature features of discrimination of the same image. This enhances classification performance in the presence of high intra-class variability.

The categorical experimental findings of the hybrid models are outlined in Tables 5 and 6, which provide the classification performance of the hybrid architectures without and with the fuzzy logic add-ons respectively.

The findings in Table 5 reveal that the fusion of feature representations of various CNN models brings about significant gains over single models. Among the pairwise, hybrid models, it was found that the best performance was DenseNet201 with ResNet50 model with the accuracy of 0.8788, overperforming the DenseNet201 with EfficientNetV2B2 model with 0.8485 and EfficientNetV2B2 with ResNet50 model with 0.7652 accuracy. The somewhat poor performance of EfficientNetV2B2 with ResNet50 combination could suggest that the features extracted by these two networks contain less complementary information than other combinations. The greatest improvement is noted in triple hybrid architecture, which combines DenseNet201, EfficientNetV2B2 and ResNet50. This model was the best in terms of overall performance with an accuracy of 0.9061 as well as high values of both the precision and recall of both the benign and malignant classes. Specifically, the model was found to have a precision of 0.9543 on benign lesions and a recall of 0.95 on malignant lesions, which is why it can be concluded that the combination of feature representations of multiple CNN architectures increases the discriminative ability of the model so that it could cover a wider range of visual patterns in dermoscopic images.

Table 6 summarizes the classification performance of the hybrid CNN models following the implementation of the decision enhancement module based on fuzzy logic.

Following the application of fuzzy logic, the accuracy of DenseNet201 with EfficientNetV2B2 hybrid model rose to 0.86 compared to 0.8485 whereas the accuracy of DenseNet201 with ResNet50 hybrid model rose to 0.89 compared to 0.8788. Likewise, EfficientNetV2B2 with ResNet50 model also experienced a limited increase in the performance of 0.7652 to 0.77. The triple hybrid model was again the best performed of all the architectures. This model achieved a precision of about 0.91 with the use of fuzzy logic. Moreover, the recall values of the malignant class in most of the models improved slightly with the application of fuzzy logic, indicating that the fuzzy decision layer is also relevant in reducing false negatives, which is particularly relevant in the medical diagnosis activity like skin cancer detection.

Comparative analysis of the hybrid CNN models in their original form and after the application of the fuzzy logic improvement are described in Figure 2 that displays the values of the Macro F1-score of each hybrid architecture. The Macro F1-score measure was compared due to the fact that it gives a balanced measure of classification performance with regard to benign and malignant classes.

The triple hybrid model was the most successful as it utilized the merits of efficient feature reuse of DenseNet201, the stability of the deep network of ResNet50, and the accuracy-speed tradeoff of EfficientNetV2B2. This combination allowed extracting complementary optical features of dermatoscopy images. So, the high intra-class variation associated with dermoscopic images can be overcame by fusion features and allows the model to reproduce a wider span of discriminative patterns.

The confusion matrix of final triple hybrid model is presented in Figure 3. The confusion matrix shows that the proposed model had a high classification performance with 315 benign cases and 285 malignant cases being classified correctly. Still, there were not many cases of misclassification, in particular, 45 benign cases were identified as malignant and 15 malignant cases as benign, which indicates that false positives were somewhat more probable than false negatives. This action corresponds with the major goal of the proposed hybrid and fuzzy structure, which is to reduce the misclassification of malignant cases.

4.3 Comparison with Previous Studies

In another attempt to analyze the efficacy of the suggested approach, Table 7 provides a comparison of the proposed approach with several current research done on dermoscopic skin cancer classification. It compares the datasets, methods, and reported performance measures of the most appropriate methods, such as transfer learning models, ensemble techniques, and hybrid deep learning models.

The contrast outlined in Table 7 shows that there are many significant points in the analysis of the recent designs of automated skin cancer classification. The majority of studies conducted in the past are based on one deep learning architecture or standard ensemble methods, including DenseNet, ResNet, or EfficientNetV2B2 models with the usage of traditional machine learning classifiers. Even though these methods have been reported to perform well, their accuracy of classification usually lie between 0.85 and 0.91 indicating the challenge that dermoscopic images are difficult to classify owing to the complex nature of the visual images of skin lesions. Moreover, a few studies strive to improve the performance of models with data augmentation methods, hand-crafted feature extraction methods, and feature selection methods, and some use transfer learning methods or uncertainty estimation methods to promote the reliability of the predictions. Nonetheless, few studies explore feature-wise fusion of two or more convolutional neural networks, which are capable of dermoscopic lesion patterns representations that are complementary. The other shortcoming that is evident in most of the existing methods is that the eventual classification decision is usually obtained out of the crude outputs of the deep learning models without considering the sophisticated decision refinement mechanisms. This limitation can lower the reliability of predictions in situations where the visual patterns of the lesions are subject to visual ambiguity or are very similar. Conversely, the suggested framework incorporates various CNN models via feature fusion and applies the decision-enhancing platform of fuzzy logic, which enhances the performance and stability of the classification procedure. The design will allow the model to be more responsive to the problem of intra-class variability of dermoscopic images, in which lesions of the same type can have considerable color, shape, and texture dissimilarities. The experiment has shown that the proposed approach provides a better accuracy of 0.91 and a higher F1-score than the analyzed studies, which can be regarded as a better balance between accuracy and recall in terms of classification. These findings validate the claim that multi-network feature fusion and fuzzy decision refinement integration helps achieve more credible skin cancer classification.

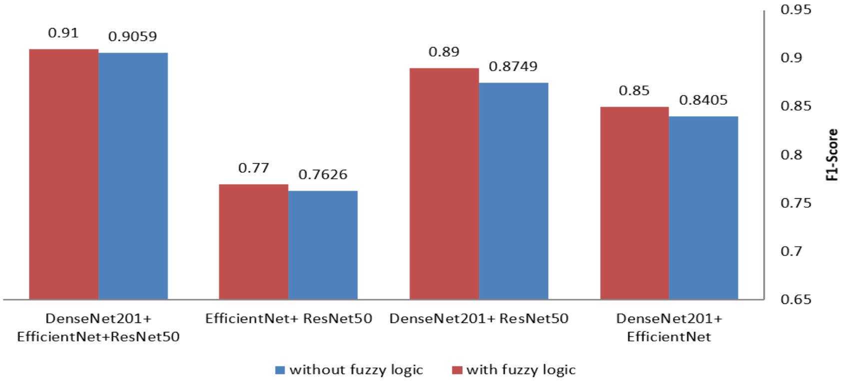

Figure 2: Macro F1-score comparison of the proposed hybrid CNN architectures with and without fuzzy logic decision enhancement.

Table 5: Classification performance of the proposed hybrid CNN architectures (two-network and three-network fusion) without fuzzy logic decision enhancement.

Class	Evaluation measure	DenseNet201 + Efficient NetV2B2	Dense Net201 + ResNet50	Efficient NetV2B2 + ResNet50	DenseNet201 + Efficient NetV2B2 + ResNet50
Benign (0)	Precision	0.80	0.86	0.88	0.95
	Recall	0.98	0.96	0.66	0.88
	F1-score	0.88	0.91	0.76	0.91
Malignant (1)	Precision	0.97	0.94	0.69	0.86
	Recall	0.71	0.82	0.90	0.95
	F1-score	0.82	0.87	0.78	0.90
Macro average	Precision	0.88	0.90	0.79	0.91
	Recall	0.84	0.89	0.78	0.91
	F1-score	0.85	0.89	0.77	0.91
Weighted average	Precision	0.88	0.90	0.80	0.91
	Recall	0.86	0.89	0.77	0.91
	F1-score	0.85	0.89	0.77	0.91
Accuracy		0.86	0.89	0.77	0.91

Table 6: Performance comparison of the proposed hybrid CNN models after applying fuzzy logic–based decision enhancement.

Class	Evaluation measure	DenseNet201 + Efficient NetV2B2	DenseNet201 + ResNet50	Efficient NetV2B2 + ResNet50	DenseNet201 + Efficient NetV2B2 + ResNet50
Benign (0)	Precision	0.7902	0.8365	0.9437	0.9543
	Recall	0.9833	0.9667	0.6056	0.8694
	F1-score	0.8762	0.8969	0.7377	0.9099
Malignant (1)	Precision	0.9717	0.9508	0.6690	0.8584
	Recall	0.6867	0.7733	0.9567	0.9500
	F1-score	0.8047	0.8529	0.7874	0.9019
Macro average	Precision	0.8809	0.8937	0.8064	0.9064
	Recall	0.8350	0.8700	0.7811	0.9097
	F1-score	0.8405	0.8749	0.7626	0.9059
Weighted average	Precision	0.8727	0.8885	0.8188	0.9107
	Recall	0.8485	0.8788	0.7652	0.9061
	F1-score	0.8437	0.8769	0.7603	0.9063
Accuracy		0.8485	0.8788	0.7652	0.9061

Table 7: Comparison between the proposed method and recent skin cancer classification studies using dermoscopic images.

Study	Dataset	Method	Key Technique	Best Performance
Hamsalekha et al. (2024) [15]	Custom dermoscopic dataset (2000 images)	DenseNet-121 (Transfer Learning)	Data augmentation + fine-tuning	Accuracy = 91%, F1 = 0.90
Hussein et al. (2025) [16]	Clinical dermoscopic dataset	HQCNN + BiLSTM + MobileNetV2	Hybrid quantum deep learning	Accuracy = 89.3%, F1 = 89.81%
Asgharnezhad et al. (2025) [17]	HAM10000	CLIP + ML classifiers	Uncertainty quantification (MCD, EMCD)	Accuracy 84.64%, F1 = 83.48using CLIP ViT + SVM
Gouda et al. (2022) [18]	ISIC 2018 (3533 images)	CNN + ResNet50 + InceptionV3	ESRGAN preprocessing	Accuracy = 85.8% using InceptionV3
Hosseinzadeh et al. (2024) [19]	Dermoscopic dataset	DenseNet201 + ML classifiers	Feature selection + ensemble learning	Accuracy = 87.72%
Priya & Rajeswari (2024) [20]	Dermoscopic images	CT + LBP + PSO + ML	Handcrafted features + optimization	Accuracy = 86.9%
Hasan et al. (2019) [21]	Public skin lesion dataset	CNN	Segmentation + feature extraction	Accuracy = 89.5%
Meghatria et al. (2025) [22]	ISIC 2019 & 2020	EfficientNetB0 + Ensemble	Curriculum learning	F1-score = 90.77%
Proposed Method	Dermoscopic skin lesion dataset	DenseNet201 + ResNet50 + EfficientNetV2B2 + Fuzzy Logic	Hybrid CNN fusion + fuzzy decision refinement	Accuracy = 0.91, F1 = 0.91

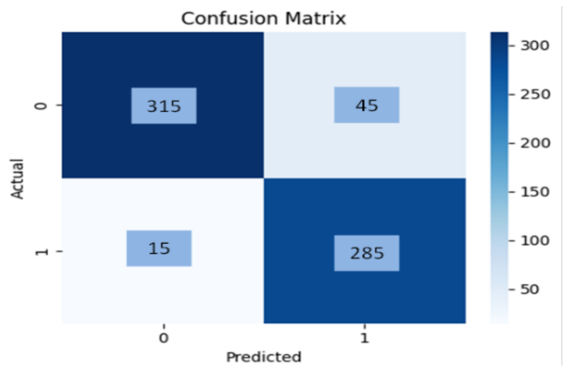

Figure 3: Confusion matrix of the triple hybrid model.

5 CONCLUSIONS

This paper suggested a hybrid deep learning model to classify skin cancer in the form of the fusion of various CNN models with a decision enhancement module of a fuzzy logic. As can be seen in the experimental results, the proposed tri-network model had the highest performance among all the models tested and had an accuracy of 0.9061, which was also enhanced to an approximate of 0.91 when fuzzy logic was applied. The results show that multi-network feature fusion improves the capability of the model to detect complementary visual patterns especially where the intra-class variation of dermoscopic images is high.

In spite of these, other issues such as variability of the datasets and image quality still exist and this could influence the generalization of the model. Although the results are promising, the study has certain limitations. The training process converged after 6 epochs, as determined experimentally (Section 3.6); additional epochs (12 and 25) were tested but did not yield further improvements in performance.

The future directions of the work include enhancing the diversity of datasets, investigating advanced frameworks, including attention-based models and Vision Transformers, and other pre-trained networks to achieve a better understanding of the model in clinical applications.

REFERENCES

[1] N. Uçar and M. F. Holick, “Illuminating the connection: cutaneous vitamin D3 synthesis and its role in skin cancer prevention,” Nutrients, vol. 17, no. 3, p. 386, 2025.

[2] E. M. Cai, P. Pathmarajah, R. Daneshjou, J. M. Ko, and A. S. Chiou, “Evaluating the appropriateness of skin cancer prevention recommendations obtained from an online chat-based artificial intelligence model,” JAAD International, vol. 18, pp. 145-147, 2025.

[3] B. Ozdemir and I. Pacal, “An innovative deep learning framework for skin cancer detection employing ConvNeXtV2 and focal self-attention mechanisms,” Results in Engineering, vol. 25, p. 103692, 2025.

[4] D. Sridhar, “Deep Learning-Based Skin Cancer Detection: A CNN Approach for Early Diagnosis,” NCAI-DS 2025, p. 33, 2025.

[5] S. Matiray and L. K. Singh, “Deep learning for early skin cancer detection: A comparative study on hybrid CNN models,” in Data, Information and Computing Science: Proceedings of the 2nd International Conference (CDICS 2024), Singapore, 6-8 December 2024, SAGE Publications, 2025, pp. 111-125.

[6] S. K. Oruganti and U. Kumar, “Early Detection of Skin Diseases Using CNN-Based Detection Systems: A Comprehensive Review,” SGS-Engineering Science, vol. 1, no. 1, 2025.

[7] S. Mavaddati, “Skin cancer classification based on a hybrid deep model and long short-term memory,” Biomedical Signal Processing and Control, vol. 100, p. 107109, 2025.

[8] M. Bhargavi and S. Balakrishna, “Hybrid approach for multi-class skin cancer classification with DCNN feature and ensemble techniques,” Engineering Research Express, vol. 7, no. 3, p. 35260, 2025.

[9] O. Yilmazer and O. K. Sahingoz, “Machine Learning Based Skin Cancer Detection Systems: A Comprehensive Survey,” in 2025 9th International Artificial Intelligence and Data Processing Symposium (IDAP), IEEE, 2025, pp. 1-8.

[10] U. Zahra, A. Manan, M. Ali, A. Ahmad, and A. Hassan, “Early Detection and Multi-Class Classification of Skin Cancer Using Deep Learning and Transformer-Based Architectures,” Spectrum Engineering Science, vol. 4, no. 2, pp. 539-556, 2026.

[11] V. Ranjan, K. Chaurasia, and J. Singh, “A comprehensive survey of skin cancer identification using deep learning,” Recent Advances in Electrical and Electronic Engineering, vol. 18, no. 9, pp. 1394-1412, 2025.

[12] M. Haque, K. Khatun, G. Rohman, A. A. Rakib, and M. Billah, “A deep learning framework for multi-class skin cancer diagnosis: Addressing class imbalance and performance optimization through convolutional neural network architectures,” European Journal of Computer Science and Information Technology, vol. 13, no. 1, pp. 63-71, 2025.

[13] I. Ashraf, M. Ullah, M. F. Nadeem, and M. N. Noor, “Coreset selection based on intra-class diversity,” arXiv preprint arXiv:2509.21380, 2025.

[14] P. Natha, S. P. Tera, R. Chinthaginjala, S. O. Rab, C. V. Narasimhulu, and T. H. Kim, “Boosting skin cancer diagnosis accuracy with ensemble approach,” Scientific Reports, vol. 15, no. 1, p. 1290, 2025.

[15] Hamsalekha R., "Fusion-Based Deep Learning Approach for Skin Cancer Detection Using DenseNet-121 and Transfer Learning," Journal of Electrical Systems, vol. 20, no. 11s, pp. 3182-3188, 2024, doi: 10.52783/jes.8044.

[16] A. A. Hussein, A. M. Montaser, and H. A. Elsayed, "Skin cancer image classification using hybrid quantum deep learning model with BiLSTM and MobileNetV2," Quantum Machine Intelligence, vol. 7, no. 2, pp. 1-28, 2025, doi: 10.1007/s42484-025-00288-y.

[17] H. Asgharnezhad, P. Tabarisaadi, A. Khosravi, R. Alizadehsani, and U. R. Acharya, "A Quad-Step Approach to Uncertainty-Aware Deep Learning for Skin Cancer Classification," 2025, [Online]. Available: http://arxiv.org/abs/2506.10302.

[18] W. Gouda, N. U. Sama, G. Al-Waakid, M. Humayun, and N. Z. Jhanjhi, "Detection of skin cancer based on skin lesion images using deep learning," in Healthcare, MDPI, 2022, p. 1183.

[19] M. Hosseinzadeh et al., "A model for skin cancer using combination of ensemble learning and deep learning," PLOS ONE, vol. 19, no. 5, pp. 1-19, 2024, doi: 10.1371/journal.pone.0301275.

[20] P. Natha and P. R. Rajeswari, "Skin Cancer Detection using Machine Learning Classification Models," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 6s, pp. 139-145, 2024, doi: 10.20944/preprints202311.0248.v1.

[21] M. Hasan, S. Das Barman, S. Islam, and A. W. Reza, "Skin cancer detection using convolutional neural network," ACM International Conference Proceeding Series, pp. 254-258, 2019, doi: 10.1145/3330482.3330525.

[22] R. Meghatria, D. Gaceb, F. Touazi, T. Boukert, and S. Ben Aidrene, "Enhancing Skin Cancer Detection Using Curriculum Learning in Ensemble Deep Learning Context," vol. 10, 2026.

[23] I. Bakurov, M. Buzzelli, R. Schettini, M. Castelli, and L. Vanneschi, "Structural similarity index (SSIM) revisited: A data-driven approach," Expert Systems with Applications, vol. 189, p. 116087, 2022.

[24] U. Gurav, R. Deshpande, and R. Mangulkar, "Soil-Type Detection: Comparative Analysis of VGG, ResNet-50, MobileNet, InceResNetV2, EfficientNet, DenseNet," ICT Applications and Social Interfaces, Proceedings of ICTCS 2024, vol. 1323, p. 11, 2025.

[25] Z. C. Oleiwi, E. N. AlShemmary, and S. Al-Augby, "Multi-label classification technique of chest X-rays image based cardiomegaly disease prediction," in 2023 International Conference on Information Technology, Applied Mathematics and Statistics (ICITAMS), IEEE, 2023, pp. 134-139.

[26] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," Journal of Big Data, vol. 9, no. 1, p. 102, 2022.

[27] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," Expert Systems with Applications, vol. 242, p. 122807, 2024.