REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

VIT-SKINNET: A NOVEL VISION TRANSFORMER-BASED SHUFFLENETV3 MODEL FOR EFFICIENT SKIN CANCER DETECTION WITH SEGMENTATION MECHANISM

Abdulmohaimen Ibrahim Khaleel AL-GBURI

Master Thesis

Supervisor

Assoc. Prof. Dr. Sefer KURNAZ

İstanbul, 2024

VIT-SKINNET: A NOVEL VISION TRANSFORMER-BASED SHUFFLENETV3 MODEL FOR EFFICIENT SKIN CANCER DETECTION WITH SEGMENTATION MECHANISM

Abdulmohaimen Ibrahim Khaleel AL-GBURI

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

2024

The thesis titled VIT-SKINNET: A NOVEL VISION TRANSFORMER-BASED SHUFFLENETV3 MODEL FOR EFFICIENT SKIN CANCER DETECTION WITH SEGMENTATION MECHANISM**,** prepared by ABDULMOHAIMEN IBRAHIM KHALEEL AL-GBURI and submitted on 10/06/2024 has been accepted unanimously for the degree of Master of Science in Electrical and Computer Engineering

_____

Assoc. Prof. Dr. Sefer KURNAZ

Supervisor

Thesis Defense Committee Members:

Assoc. Prof. Dr. Sefer KURNAZ	Department of Computer Engineering, Altınbaş University	_____
Assoc. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altınbaş University	_____
Asst. Prof. Dr. Serdar KARGIN	Department of Biomedical Engineering, İstanbul Arel University	_____

I hereby declare that this thesis meets all format and submission requirements of a master's thesis.

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Abdulmohaimen Ibrahim Khaleel AL-GBURI

Signature

DEDICATION

I dedicate this research to my supervisor, Assoc. Sefer Kurnaz. I would also like to extend my thanks and gratitude to my father and mother for their continuous support throughout my study period. I also thank all my family members, friends, and colleagues during my study period, and everyone who encouraged me, supported me, and stood by me.

PREFACE

This research was conducted in accordance with the requirements for the master's degree in computer engineering at Altinbaş University under the supervision of Assoc. Prof. Dr. Sefer Kurnaz.

ABSTRACT

VIT-SKINNET: A NOVEL VISION TRANSFORMER-BASED SHUFFLENETV3 MODEL FOR EFFICIENT SKIN CANCER DETECTION WITH SEGMENTATION MECHANISM

AL-GBURI, Abdulmohaimen Ibrahim KHALEEL

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Dr. Sefer KURNAZ

Date: 06/ 2024

Pages: 69

Melanoma, a deadly type of skin cancer, claims hundreds of lives every year. Skin cancer typically affects areas of skin that are exposed to sunlight on a frequent basis, such as the "legs, face, arms, and neck". By visually examining lesions with pigment on the skin, melanoma can be identified early and treated with a straightforward method of removal of the malignant cells. However, the examination of the skin by the naked eye alone has a restricted and inconsistent accuracy owing to the scarcity of dermatologists. This also leads the patients to undergo multiple biopsies and thus hampers the course of treatment. The automatic identification of lesions from the dermoscopy images involves several obstacles due to the intricate lesion characteristics and as a result of the detection backdrop. There is a dearth of research on major intra-class variations and inter-class similarities of lesion characteristics, and the prior solutions primarily concentrate on employing larger and more complicated systems for detecting the presence of skin cancer with much-enhanced detection accuracy. In order to address various issues with the traditional methods, it is therefore increasingly crucial to create a successful structure for the identification of skin cancer through the use of deep learning approaches. The implemented skin cancer detection and classification model has three crucial procedures to perform. The collection of dermoscopy images, the segmentation process for segmenting the images, and the detection phase are the three main stages of the implemented skin cancer detection system that is put into practice. First, the benchmark images are utilized to provide the images that are needed for the tests.

Once the images are gathered, then the segmentation phase is executed. The segmentation step receives the gathered images as its input. Then, the Residual DenseUNet++ (ResDenseUNet++) is used to carry out an effective segmentation. At this point, the resultant segmented image is provided by the developed ResDenseUNet++, which is then considered for the later stage'sinputs. Additionally, the segmented image is then given to the feature extraction process by which the gradient filter image and the texture pattern image are obtained. Additionally, the obtained image results are then inputted for the phase of skin cancer detection. During the detection phase, the newly developed ShufflenetV3 based on Vision Transformer (ViT-ShufflenetV3) is implemented and used to effectively recognize skin cancer. In many experimental validations, the recommended skin cancer identification system has ensured a more precise classification outcome regarding skin cancer than other traditional models.

KEYWORDS: Skinnet**,** DenseUNet++, Vision Transformer**,** Segmentation, Convolutional Neural Network (CNN)**.**

ÖZET

VIT-SKINNET: SEGMENTASYON MEKANİZMASI İLE ETKİLİ CİLT KANSERİNİN TESPİTİ İÇİN YENİ BİR VİZYON TRANSFORMATÖR TABANLI SHUFFLENETV3 MODELİ

AL-GBURI, Abdulmohaimen Ibrahim Khaleel

Yüksek Lisans, Elektrik ve Bilgisayar Mühendisliği, Altınbaş Üniversitesi

Danışman: Doç.Dr. Sefer KURNAZ

Tarih: 06 / 2024

Sayfa: 69

Ölümcül bir cilt kanseri türü olan melanom, her yıl yüzlerce kişinin ölümüne neden oluyor. Cilt kanseri tipik olarak "bacaklar, yüz, kollar ve boyun" gibi cildin sıklıkla güneş ışığına maruz kalan bölgelerini etkiler. Derideki pigmentli lezyonların görsel olarak incelenmesiyle melanom erken tespit edilebilir ve kötü huylu hücrelerin çıkarılması gibi basit bir yöntemle tedavi edilebilir. Ancak dermatoloğun azlığı nedeniyle derinin yalnızca çıplak gözle incelenmesi sınırlı ve tutarsız bir doğruluğa sahiptir. Bu durum aynı zamanda hastaların birden fazla biyopsi almasına neden olmakta ve tedavi sürecini aksatmaktadır. Lezyonların dermoskopi görüntülerinden otomatik olarak tanımlanması, karmaşık lezyon özelliklerinden ve tespit zemininin bir sonucu olarak çeşitli engeller içerir. Lezyon özelliklerinin ana sınıf içi varyasyonları ve sınıflar arası benzerlikleri üzerine araştırma eksikliği vardır ve önceki çözümler öncelikle, çok daha gelişmiş tespit doğruluğu ile cilt kanserinin varlığını tespit etmek için daha büyük ve daha karmaşık sistemlerin kullanılmasına odaklanmaktadır. Çeşitli sorunları geleneksel yöntemlerle ele almak amacıyla, derin öğrenme yaklaşımlarını kullanarak cilt kanserinin tanımlanmasına yönelik başarılı bir yapı oluşturmak giderek daha önemli hale geliyor. Uygulanan cilt kanseri tespit ve sınıflandırma modelinin gerçekleştirilmesi gereken üç önemli prosedür vardır. Dermoskopi görüntülerinin toplanması, görüntülerin segmentlere ayrılması için yapılan segmentasyon süreci ve tespit aşaması, uygulamaya konulan cilt kanseri tespit sisteminin üç ana aşamasıdır. İlk olarak, testler için gerekli görüntüleri sağlamak amacıyla kıyaslama görüntüleri kullanılır.

Görüntüler toplandıktan sonra segmentasyon aşaması gerçekleştirilir. Segmentasyon adımı, toplanan görüntüleri girdi olarak alır. Daha sonra etkili bir segmentasyon gerçekleştirmek için Residual DenseUNet++ (ResDenseUNet++) kullanılır. Bu noktada, elde edilen parçalı görüntü, geliştirilen ResDenseUNet++ tarafından sağlanmakta ve daha sonraki aşamanın girdileri için dikkate alınmaktadır. Ek olarak, parçalı görüntü daha sonra özellik çıkarma işlemine tabi tutularak degrade filtre görüntüsü ve doku deseni görüntüsü elde edilir. Ek olarak, elde edilen görüntü sonuçları daha sonra cilt kanseri tespit aşaması için girilir. Tespit aşamasında, Vision Transformer'ı (ViT-ShufflenetV3) temel alan yeni geliştirilen ShufflenetV3 uygulanır ve cilt kanserini etkili bir şekilde tanımak için kullanılır. Birçok deneysel doğrulamada, önerilen cilt kanseri tanımlama sistemi, diğer geleneksel modellere göre cilt kanserine ilişkin daha kesin bir sınıflandırma sonucu sağlamıştır.

Anahtar Kelimeler: Skinnet, DenseUNet++, Vision Transformer, Segmentasyon, Evrişimsel Sinir Ağı.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

1. INTRODUCTION

Among the various common forms of cancer that exist in the current decade, skin cancer is a general one. Given that the skin is the biggest tissue in the body, it makes sense to regard skin cancer as the most prevalent kind of cancer in people. In accordance with the cancer stats provided by the "American Cancer Society (ACS)", skin cancer is considered one of the cancer types that lead to a significant rise in the death rate. Patients who have skin cancer also have a higher death rate. This variety of carcinoma has a particularly high death rate, which is continually rising with every occurrence. Thankfully, there is an increased likelihood of survival if the condition is discovered and diagnosed quickly in the very beginning stages [9]. There are two types of skin cancer, namely, "non-melanoma and malignant melanoma". One of the most dangerous, uncommon, and fatal kinds of skin cancer is regarded as melanoma. The ACS reports that although melanoma skin cancer accounts for only 1% of all occurrences, its fatality rate are greater than other type of skin cancer. Although just 4% of people are estimated to be affected, malignant melanoma accounts for 75% of skin cancer-related deaths [10]. It is one of the most serious and deadliest types of cancer. Cells that produce "melanocytes", which are found throughout the body, are what produce melanoma. Melanocytes are the cells from which the melanoma arises from. It begins when normal melanocytes start to proliferate uncontrollably, giving rise to a malignant tumor. Any part of the human body is susceptible to being affected by melanoma. Typically, it shows up on the lips, face, hands, neck, and other area that are highly exposed to sunlight. There are several kinds of melanoma skin cancer, including "lentigo maligna, acral lentiginous, superficial spreading melanoma, and nodular melanoma". Nonmelanoma classifications, including "Sebaceous Gland Carcinoma (SGC), Basal Cell Carcinoma (BCC), and Squamous Cell Carcinoma (SCC)" encompass the majority of cancer cases. The central and topmost layers of the epidermis are SCC, BCC, and SGC are the regions from which this type of cancer originate from. The propensity of these cancer cells to spread across different body parts is minimal.

Compared to melanoma cancers, non-melanoma cancers are easier to treat. A timely diagnosis is therefore essential to the successful treatment of skin cancer. The biopsy procedure is typically used by doctors to identify skin cancer when it is suspected. In order to assess if a suspicious skin lesion is malignant or not, a sample is taken during this operation. This is a gradual and painful procedure that takes a long time. Cancers of the

melanoma type are incurable and can be treated only if discovered at a very beginning stage. If not, they spread to different regions of the body and cause a terrible death for the patient affected by it. When melanoma becomes apparent or confirmed in the very beginning phases and is treated promptly, it can be healed. However, in the event of melanoma when discovered in its later stages could grow more deeply into the skin and have an impact on different regions of the body as well. In such cases, the treatment may become more challenging [11]. "Dermoscopy" is a commonly used non-traumatic skin imaging tool that is widely utilized in melanoma detection [12]. Despite the fact that melanoma can be detected more accurately with the aid of Dermoscopy in an effective manner than when compared with other conventional techniques used for the detection of melanoma such as the one with no additional inspection, dermatologists' knowledge and expertise are key factors in determining the diagnostic accuracy of melanoma. The accuracy in diagnosing melanoma can only be reached by dermatologists, however, consistency in making keen observations for generating accurate results and the variation in perception among distinct doctors' diagnoses is also different making the accuracy of detection lower and the overall detection a complicated and fluctuating task [13]. Thus, it is extremely important from the practical point of view to use the benefits of artificial intelligence to help physicians with non-contact and automated diagnosis of skin cancer [14].

Computer-based technology makes it possible to diagnose skin cancer symptoms quickly, affordably, and comfortably without much pain. Multiple non-invasive procedures are offered to investigate the signs of skin cancer, to determine whether or not they are indicative of melanoma or not. The advancements of machine learning and artificial intelligence have opened up novel opportunities for assistance diagnostics in the healthcare and biomedical domains [15]. Traditionally, for the detection of skin cancer, the images of skin diseases are considered for the process of characteristics extraction, and the characteristics that are being extracted are subsequently classified, in order to diagnose the skin cancer [16]. Following that, a variety of classification techniques, including "Decision Trees (DT), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost)" are used to classify the crucial characteristics that have been obtained by means of handcrafted methods [17]. Machine learning algorithms are frequently used to classify only one subset of skin melanoma and are unable to generalize to a broad variety of skin disease types because of a limited amount of specified characteristics. Furthermore, it is ineffective in distinguishing different kinds of

cancer purely from physical aspects due to the diversity of skin cancers [18]. Improving the diagnosis accuracy of skin malignancies is the primary objective of employing machine learning in classifying skin cancer. Nevertheless, the accuracy attained with this approach was inferior to that of the one obtained by means of dermoscopy [19]. A few of the obstacles that machine learning must overcome include the scarcity of organized health data for pertaining the model to detect skin cancer, the lack of consistency in the gathering and the preservation of data that is not structured, and the scarcity of diversified and higher-quality data for the detection of skin cancer [20]. Owing to these difficulties, skin cancer detection uses deep learning algorithms.

In recent decades, deep learning has completely changed the field of machine learning techniques. It is regarded as the most advanced area of machine learning that deals with algorithms based on artificial neural networks. The structure and operation of the human brain served as the model for these algorithms. Many fields, including "bioinformatics, pattern recognition, and speech recognition" use deep learning techniques. When compared to other traditional machine learning techniques, deep learning systems have produced remarkable outcomes in various applications. In recent years, a variety of deep-learning techniques have been applied to computer-based skin cancer screening. When it comes to melanoma identification and segmentation, deep learning-based methods outperform the conventional hand-crafted feature-based methods. "Obtaining the image, pre-processing it, segmenting the pre-processed image, extracting the desired feature, and classifying it" are the general steps in the skin cancer detection method using deep learning models. "Deep Neural Networks (DNN)" is important for the diagnosis of skin cancer. They are made up of several linked nodes. Their structure is comparable to that of the human brain in terms of the interconnectivity of neurons. Their nodes collaborate to find solutions to specific issues. After being trained for certain tasks, neural networks function as authorities in the fields for which they were trained. Skin cancer detection systems use conventionally distinct methods of learning the disease, such as "K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Generative Adversarial Network (GAN)". Although deep-learning approaches are much more popular for classifying skin cancer in recent years, there are still a number of obstacles to overcome [21]. Deep learning models have the ability to evaluate information gathered from a large-scale of data in a quicker and more correct manner than the typical machine learning techniques, which

enables them in identifying the important features in a more efficient manner deep learning algorithms, however, can also help physicians analyze the data and examine the obtained test results more thoroughly and effectively [23]. Numerous research studies have shown that deep learning techniques are capable of diagnosing skin cancer at a similar level to that of a dermatologist. To become an extensive diagnostic system, these techniques continue to encounter numerous challenges [24]. Generally speaking, medical imaging uses CNN extensively for segmentation and classification purposes. In the task of classifying skin lesions, the CNN models performed better than a group of board-certified dermatologists [25]. The highest accuracy has been achieved by CNNs in the case of visual imaging tasks, thus making it suitable for skin cancer classification tasks.

1.1 PROBLEM STATEMENT

The main challenges of skin cancer detection techniques are they face difficulties in determining the accurate malignant versus at the beginning stage of the tumor with few etiologies and similarities such as color, shape, and border and it is difficult for dermatologists to detect the tumor cells because they are derived from Melanocytes cells. Detection and examination of skin lesions through manual examination and visual inspection are always inconvenient. Several researches were performed by the researchers in the detection of skin cancer. The advantages and disadvantages of the traditional skin cancer detection techniques are shown in Table 1. EfficientNet and Vision Transformer [1], utilize fewer parameters during high computational processes and provide a better configuration rate. Yet, it consumes more time in training the network and requires large memory space. CNN [2], provides stable outcomes during segmentation and improves diagnostic accuracy. But it requires more images for training and time-consuming tasks. MobileNet and Dense Net [3], attain improved segmentation accuracy while handling with fewer parameters. However, it is expensive for implementation and requires a large amount of computational resources. RCNN, FCM [4], is capable of handling large datasets and predicting the features of different skin diseases at the same time. Yet, it consumes more time for computation and has less generalization ability. Deep Learning Studio (DLS) [5], is capable of handling large datasets and has a high generalization ability. DNN [6], is capable of processing structured and unstructured data and it easily handles complicated and huge data. Yet, it requires more data for training the network and is costly. Generative Adversarial Network, CNN [7],

doesn't require labeled data to train the network. However, this method has a low computational rate and consumes huge resources for processing. Caps Net [8], requires less data for training the network and provides a better segmentation rate. Still, this method has high implementation complexity and is time-consuming. So, it is necessary to develop an effective framework to overcome the challenges faced by the existing system and also to improve skin cancer detection by utilizing deep learning approaches.

1.2 MOTIVATIONS

The following is the motivation that is considered for the development of this skin cancer detection model.

a. There is a very little method available to characterize the appearance of skin cancer because these changes might vary widely.
b. The manual detection of skin cancer is a complicated and time-consuming task.
c. The detection of skin cancer by manual means is prone to error and the detection outcomes are biased by the perspective of the clinical professionals.
d. Skin cancer if note detected in the early stage, may spread to other parts of the body and also causes harm to the internal organs as well.
e. When timely detection of skin cancer is made, then suitable treatment can be provided at a very early stage, thus aiding in the reduction of the mortality rate in humans.

1.6 OBJECTIVES

The objectives of the implemented skin cancer detection model are provided as follows.

a. To collect the data regarding dermoscopy images of skin cancer in order to test and train the implemented deep learning model.
b. To implement a deep learning-based lung cancer detection model in order to identify the skin lesions at the very beginning stage.
c. To utilize advanced deep learning-based segmentation techniques in order to carry out efficient skin cancer detection.
d. To analysis the performance of the executed model and t compare its results with the conventional detection techniques.

1.7 CONTRIBUTIONS

The major contributions of this synopsis work are listed in the below section.

a. To elaborately analysis the existing deep learning-based skin cancer detection model in order to analysis the challenges in existing models and to utilize the advantages of these models in developing an efficient deep learning-based skin cancer detection model.
b. To perform an advanced deep learning-based segmentation task to segment the dermoscopy images effectively in order to obtain the texture pattern images and the Gabor filter images for assisting in effective skin cancer detection tasks.
c. To make use of the deep learning model in detecting skin cancer in a much easier and faster way by utilizing advanced image processing and segmentation techniques so that to assist in the timely detection of skin cancer in patients at the very beginning stages.
d. To validate the performance offered by the executed segmentation-assisted detection model by comparing it with the functioning of the other existing segmentation approaches and conventional detection approaches in order to showcase the effectiveness of the proposed approaches.

1.8 CAUSES AND SYMPTOMS OF SKIN CANCER

Skin cancer is caused as a result of high exposure to sunlight. Areas of skin that are vulnerable to sunlight, such as the "face, scalp, ears, lips, chest, neck, and hands, arms, as well as the legs" in women, are the main sites where skin cancer occurs. However, it can also develop in places of your body that are seldom exposed to daylight, such as the space under the tips of your nails or toenails, vagina, and palms. All skin tones, particularly those with darker complexions are susceptible to skin cancer. People with darker complexions are more prone to develop melanoma on parts of their bodies like the soles of their feet and palms of their hands that are not often exposed to the sunlight. Some of the symptoms of skin cancer are given in the below section.

a. An Alteration in the existing spot's current dimensions, appearance, or colour.
b. A new area of skin that appears suddenly.
c. A bleeding or crusting sore that is not healing.
d. An area of skin that hurts or itches.
e. A new red, hard, or scaly area in the skin.
f. A shiny, reddish-coloured pimple on the skin's surface.

g. A growth that resembles a scar but lacks a clear border.
h. An inflammatory growth.
i. A growth that has a raised edge and a crusty or bleeding center.

1.9 STAGES IN SKIN CANCER

a. Skin cancer is categorized into various stages which are given as follows:
b. Stage 0 Cancer: Another name for stage 0 is cancer in situ. The term "in situ" refers to the state in which the cells are still developing. Thus, although the cells have begun to develop into cancer, they have not yet proliferated or disseminated into the skin's tissues that surround them.
c. Stage 1 Cancer: Cancer in Stage 1 is defined as having a diameter of no more than 2 cm.
d. Stage 3 Cancer: In Stage 3, the skin cancer may not have progressed to any lymph nodes yet may still be bigger than 4 cm, have expanded past the fat layers beneath the skin's surface, have invaded the area surrounding a nerve, or have caused some harm to neighbouring bones.
e. Stage 4 Cancer: This stage of the disease can indicate that it has progressed to one or more lymph nodes and can be of any size. It can mean that the cancer has expanded to bone marrow or bone, such as the bottom of the skull leading to the weakening of the bone, or that the cancer has propagated to a different and distant part of your body, like the lungs. Cancer develops through the outermost layer of one lymph node and expands to multiple lymph nodes, propagated to solely a single lymph node that is bigger than 3 cm and less than 6 cm, or propagated to lymph nodes on the opposite side of the body that is adjacent to the skin cancer, all denotes the skin cancer.

1.10 SKIN CANCER EFFECTS AND ITS PREVENTIVE MEASURES

When skin cancer affects an individual, it leads to certain harmful effects on the human body such as

a. Itching, tingling sensation, sudden pain, or numbness in the affected area.
b. Change in skin colour to pale, pearly shade, or red colour.
c. The sudden appearance of a scaly area and dry skin.
d. Failure to heal the wounds that occur on the skin.
e. A crawling-like sensation that occurs on the skin.

The occurrence of skin cancer can be prevented by means of taking the following actions.

a. Staying inside from the reach of the sunlight, thus avoiding the effects of harmful radiation.
b. Wearing cloths to cover the area of skin from the morning sunlight.
c. The utilization of water-resistant sunscreen has an SPF value higher than 30 with a broader spectrum.
d. Avoid the chances of sunburn and tanning
e. Quit smoking as it is the cause of many types of cancer.
f. Note for the changes in the skin.

1.11 ADVANTAGES OF DETECTING SKIN CANCER

As skin covers the body, it is considered to be the largest part of the human body. The skin protects the internal organs from various external factors. The skin cancer initially appears as a mole or a small-sized variation in the skin. However, if it is not detected in the earliest possible way, it may spread to distinct parts of the body. So, it is crucial to detect skin cancer effectively in the beginning stage itself. Also, if the skin cancer spreads to distinct parts of the body and penetrates deep inside the skin, then the chances of the cancer affecting the internal organs are higher leading to a higher death rate. So, to prevent the spreading of the cancer from affecting other parts of the body and to prevent it from affecting the internal organs, skin cancer has to be figured out at its initial stage. Detection of skin cancer makes the diagnosis and treatment procedure much easier and effective.

1.12 NEED FOR SEGMENTATION TECHNIQUES IN THE SKIN CANCER DETECTION PROCESS

Conventional skin cancer detection models perform segmentation procedures on the skin cancer images to make the detection procedure much easier and more effective. This makes the computation process involved in the detection model less complicated and reduces the overall detection time as well. Some of the segmentation approaches that are widely used in the segmentation of skin cancer are “Gray-Level Co-occurrence Matrix (GLCM), multi-level thresholding technique for segmentation of images, variations in CNN, and dermoscopy. Segmentation techniques have certain advantages as provided below.

a) Helps in efficient detection and classification of skin cancer.
b) Helps to extract the crucial features from the provided input image in an efficient manner.
c) Helps in analyzing the image-based data effectively.

Thus, to make skin cancer detection more effective and error-free, segmentation procedures are required in the process.

1.13 DEEP LEARNING TECHNIQUES FOR SKIN CANCER DETECTION

Deep learning models can be used to accurately detect and classify lung cancer in an effective manner. This deep learning model also aided in the identification of skin cancer in the very beginning stage with the utilization of deep features. The main focus behind the utilization of skin cancer by deep learning models is to enhance the accuracy. Some of the deep learning-based skin cancer detection model is as follows.

CNN-based skin cancer detection: The most frequently used image recognition-based deep learning model for the skin cancer detection task is CNN. CNN is used as it is capable of detecting skin cancer directly from the provided dermoscopy images. CNN typically utilizes distinct convolutional operations, followed by pooling to extract the necessary features to perform the skin cancer detection task. A huge number of hidden layers are present in the CNN model. These layers assist in the processing of the provided input images. The most basic unit present in the CNN is the convolutional layers. These convolutional perform most of the operations. After performing the pooling operation and ReLU function, the final detected images are obtained from the CNN model. Some of the variants of CNN are given as follows.

a. AlexNet
b. “Visual Geometry Group (VGG)”
c. Residual Network (ResNet)
d. Dense Network (DenseNet)
e. MobileNet

These modes can also be seen widely in skin cancer detection tasks.

AlexNet [26]: AlexNet is a variant of CNN with 3 densely connected layers and 5 convolutional layers. This model is an efficient image classification model. AlexNet is

known for its capacity in solving overfitting issues. The overfitting issues are solved in the AlexNet model with the utilization of the dropout layer.

VGG [27]: On the basis of the number of convolutional layers, the variation in the VGG can be made. The frequently utilized VGG models for image processing tasks are the VGG-19 and VGG-16 models. VGG models have 16/ 19 convolutional layers, followed by densely linked layers, maximum pooling layers, and an output layer.

ResNet [28]: The issues present in the deep learning model when the number of layers in the networks gets increased, vanishing gradient issues, can be resolved with the implementation of the ResNet model. In the ResNet model, the skip connections are introduced for solving these issues. Skip connections aid in making interactions between various layers in the networks. Thus, effective detection of skin cancers is made possible by means of the ResNet model.

DenseNet [29]: The DenseNet model utilizes various layers whose inputs are the output from the prior layers. The previous layer's feature maps are given as the input to the next layer. This aids in the propagation of the features from one layer to another thus minimizing the parameters required for the model.

MobileNet [30]: In order to make a light weight model for detecting skin cancer, MobileNet is used as it is portable in nature. MobileNet model utilizes the depth-wise separable and convolutional operations.

Random Forest [31]: An ensemble model that utilizes a decision tree for performing various regression and classification tasks is called a random forest model. Then averaging is performed on the final outputs from all the decision trees to get the finally classified output.

2. LITERATURE SURVEY

In 2021, Duong et al. [1], have utilized state-of-the-art computer vision and machine learning methods to design a workable method for detecting tuberculosis using "Chest X-Ray (CXR)" images. A conceptual system using "updated EfficientNet, updated classic ViT, and updated Hybrid EfficientNet together with ViT" were utilized as the primary classification engines, all based on DNN. Furthermore, a variety of augmentation approaches were then utilized to strengthen the process of learning. The suggested method was assessed using a sizable dataset that was assembled by combining many publicly available datasets. 80%, 10%, and 10% were taken into consideration for dividing the actual dataset each for "training, validation, and testing" groups, individually. After splitting the dataset, the model was trained and was then validated to prove its effectiveness. The experimental analysis proved the enhanced performance offered by the implemented deep learning-based tuberculosis detection model.

In 2019, Dascalu and David [2], have implemented the "Skin Magnifier with Polarized Light (SMP)" approach for sonification for the purpose of ascertaining the influence of image quality on the diagnosis accuracy of skin cancer detection. After being captured by SMP, the gathered dermoscopy images were sonified and subjected to initial processing with the aid of a deep learning method. An additional supplementary deep-learning approach was used to examine the audio output. "Sensitivity and specificity" were processed further by utilizing an additional metric as given by F2-score which was obtained by giving sensitivity twice the weight as much as the positive prediction values in order to study the key findings associated with the SMP. A total of seventy-three patients who met the requirements for participation were sent for a biopsy test. An enhanced dermoscopy which when used to diagnose an identical set of lesions in patients has produced similar results including "59.9% positive predictive scores, 89.5% F2-score sensitivity, and 57.8% specificity, respectively on experimental validation.

In 2020, Wei et al. [3], have generated "fine-grained classification principle-based feature discrimination" for identifying skin cancer with a lightweight structure. Two common feature extraction modules from a feature discrimination model along with a cancer classification model were included in the proposed structure. First, the recognition model has extracted features by means of a lightweight CNN model, in which positive as well as negative samples were received as two sets of training data. The two recognition models,

including the classification model and the feature discrimination model, were then trained simultaneously using a pair of feature vectors output from the feature extraction module, and the model fusion strategy was used to further enhance the model's performance. The suggested recognition method extracted more discriminative cancer characteristics and enhanced the model's recognition performance with a limited number of parameter requirements. Furthermore, a compact semantic segmentation system for the cancer-affected region of a Dermoscopy image using the feature extraction unit of the suggested detection model using UNet structure along with migration training strategy was used. This model allowed for high-precision segmentation of cancer-affected regions to be achieved in an automatic manner without the need for complex image preprocessing operations. The efficacy of this methodology was evaluated by means of feature visualization as well as comparative analysis with the aid of numerous experiments. Experimental results showed that the suggested approach outperformed the state-of-the-art deep learning-based skin cancer detection method on the "ISBI 2016 skin lesion analysis towards the melanoma detection challenge dataset".

In 2019, Nida et al. [4], have addressed the difficulties in the automatic segmentation of the Melanoma zone created by means of a deep learning technique with the aid of utilizing the dermoscopic images. The segmentation was performed on the dermoscopic images. The numerous impacted regions were accurately identified by means of a "deep Region-Based Convolutional Neural Network (RCNN)" as bounding boxes for making the overall localization process much easier using the "Fuzzy C-mean (FCM)" clustering technique. The three major procedures that were considered while developing this model were "refinement of skin, melanoma region localization, and melanoma segmentation". The suggested approach was used by executing it in the "International Symposium on Biomedical Images (ISBI) benchmark dataset, ISIC-2016", which contained a total of 376 images dedicated to testing and 900 images dedicated to training these dermatological images for the detection of melanoma. Melanoma segmentation performance was assessed using a range of quantitative metrics. Experimental examination against cutting-edge techniques has shown that the outcomes obtained from the recommended model were much superior to that of several other existing detection and segmentation approaches.

In 2020, Kadampur and Riyaee [5], have generated a model based on deep-learning techniques to identify skin cancer by classifying images that consisted of dermal cells.

Models with better skin cancer prediction were created using a model-driven structure in the cloud environment that relies on deep learning techniques for its fundamental operations. The work demonstrated how to create models and use them to categorize images of dermal cells. Using benchmark datasets, the implemented deep learning systems were trained and evaluated. The "Area under the Curve (AUC)" was regarded as the validation metric. When evaluated, the implemented model has shown an AUC of around 99.77%. In order to anticipate skin cancer in an accurate manner, practitioners swiftly developed deep learning frameworks using the "model-driven" approaches.

In 2023, Venugopal et al. [6], have generated and presented a DNN model for the diagnosis of skin cancer that has better learning performance by adjusting its training in an effective manner. The DNN systems were trained using a knowledge base that was created by merging various dermoscopic datasets. To expedite the training of the suggested model on a small-scale training dataset, the methods of "fine-tuning and transfer learning" were used. The model's performance was improved by using the data augmentation strategies. There were around 58,032 enhanced dermoscopic images that were utilized in total. To carry out the task of binary classification for skin cancer detection, the final result from the multilayered structure was combined. Investigations were conducted to examine the trained models' performance in binary and multiclass classification tasks. According to various performance measures, the DNN model with enhanced EfficientNetV2-M has performed better than the most advanced multi-class classification model that was developed with the utilization of deep learning techniques.

In 2023, Lembhe et al. [7], have designed a computerized dermatological cancer screening procedure by utilizing advanced image processing techniques and machine learning methods. Using visual LR images, "Image Super-Resolution (ISR)" algorithms were generated with an extremely higher quality of images or sequences. To increase the accuracy of the implemented CNN model, a deep learning method on the ISR was employed. The Keras backend was utilized in the development of this model, and the layers of the neural network that were employed for training were modified to verify the working of the implemented technique. The model was constructed using freely available datasets from the archived ISIC data. Experimental results revealed the enhanced efficacy of the implemented technique on skin cancer detection tasks.

In 2021, Adla et al. [8], have produced a model named the DLCAL-SLDC, an end-to-end deep learning-based model with class attention layers with a computer-assisted skin lesion classification model. The DLCAL-SLDC model sought to identify and categorize the various Using dermoscopic images, the DLCAL-SLDC model aimed to identify and categorize the various forms of skin cancer. Hair removal using a dull razor technique and noise reduction using average median filtering were used during the image pre-processing procedure. The dermoscopic images were then analyzed using a segmentation technique based on "tsallisentropy" to identify the afflicted lesion regions. Additionally, a DLCAL-based tool for extracting the feature was employed in conjunction with Adagrad and CAL optimizer to obtain characteristics from the segmented lesions utilizing Capsule Network (Caps Net). In order to cover the class dependencies and successfully bridge the CapsNet for additional processing, the CAL layer that was included in the CapsNet was designed in obtaining the selective class-specific characteristics. At the end of the "Convolutional Sparse Auto Encoder (CSAE)" denoted as the SSO-CSAE model, which was based on the "Swallow Swarm Optimization (SSO)" algorithm, the classification task was performed. The benchmark ISIC dataset was used to validate the suggested DLCAL-SLDC approach. In comparison to the previous approaches, the suggested framework has achieved impressive results with an accuracy of 98.50%, sensitivity of 94.5%, and specificity of 99.1% when contrasted with a range of metrics.

Table 2.1: Features and Challenges in Existing Deep Learning-based Skin Cancer Classification Model.

Author [citation]	**Methodology**	**Features**	**Challenges**
Duong *et al.* [1]	EfficientNetand ViT	• This method utilizes fewer parameters during a high computational process. • This method provides a better configuration rate.	• It consumes more time to train the network. • It requires large memory space.
Dascalu and David [2]	CNN	• This method provides stable outcomes during segmentation. • This method improves diagnostic accuracy.	• It requires more images for training. • It is time-consuming.

Table 2.1: Features and Challenges in Existing Deep Learning-based Skin Cancer Classification Model "Table Continued".

Wei *et al.*[3]	MobileNetand DenseNet	• This method attains improved segmentation accuracy while handling with fewer parameters.	• It is expensive for implementation. • It requires a large amount of computational resources.
Nida *et al.* [4]	RCNN and FCM	• This method is capable of handling large datasets. • This method predicts the features of different skin diseases at the same time.	• It consumes more time for computation. • It has less generalization ability.
KadampurandRiyaee [5]	Deep Learning Studio (DLS)	• This method is capable of handling large datasets. • This method has high generalization ability.	• This method lacks interpretability among the data. • It has a high computational cost.
Venugopal *et al.* [6]	DNN	• This method is capable of processing structured and unstructured data. • It easily handles complicated and huge data.	• It requires more data for training the network. • It is costly.
Lembhe*et al.* [7]	GAN and CNN	• This method doesn't require labeled data to train the network.	• This method has a low computational rate. • It consumes huge resources for processing.
Adla *et al.* [8]	CapsNet	• This method requires less data for training the network. • This method provides a better segmentation rate.	• This method has high implementation complexity • It is time-consuming.

3. METHODOLOGY

Out of several varieties of diseases, one of the most deadly diseases that have threatened the lives of people in the past few decades is cancer. Skin cancer is one of the prevalent types of cancer that affects women in general. Taking into consideration several kinds of cancer classification techniques, skin cancer classification and detection appear to pose a significant challenge. The success of the predictive algorithm depends on ideal features that are being extracted, which a significant problem in machine is learning as well as data mining-based skin cancer diagnosis and classification models. Manual extraction of these ideal features necessitates a domain-level understanding of the processing of images and a thorough understanding of skin conditions. Consequently, deep learning-based techniques are developed to support improved performance in the classification of skin cancer. Since manual identification and detection of skin cancer is laborious and inaccurate, automatic deep-learning models are generated and are now prominently in use. Yet, as a result of their higher level of complexity, considerable computational overhead, and inadequate representation of features, the current methods are capable of generating less accurate detection outcomes. Therefore, in this work, a new deep learning model for identifying skin cancer is suggested to overcome the current limitations in the existing models. The pictorial representation of the implemented skin cancer detection model is depicted in Figure 1.

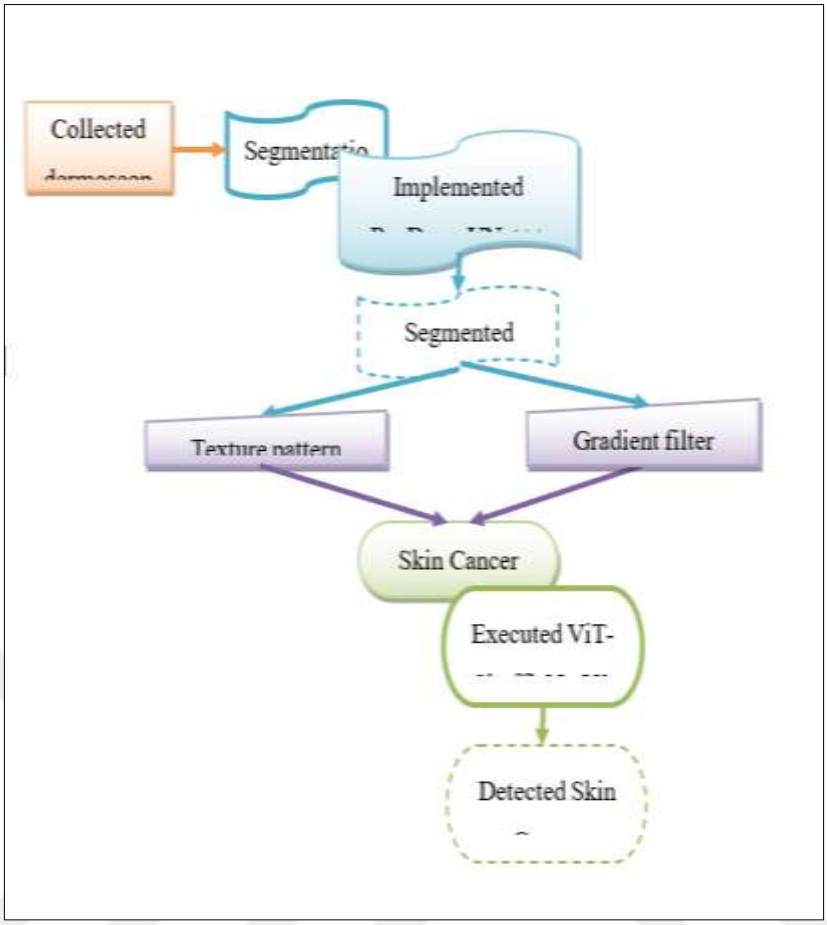


Figure 3.1: Pictorial Representation of the Implemented Skin Cancer Detection Model.

A deep learning-based skin cancer detection framework is implemented in this work to assist the radiologists in identifying the skin cancer at the very beginning stage. This model is aimed to resolve the disadvantages associated with the traditional deep learning-based skin cancer classification model. The essential dermoscopy images are gathered from online platforms. The collected images are fed as input to the segmentation stage. The segmentation model utilized in this work is the ResDenseUNet++ model. The segmentation of the dermoscopy image by utilizing the ResDenseunet++ model is then carried out. From the segmentation model, the texture pattern image and gradient filter image are obtained. These image outputs are then given as the input to the detection model. The detection model used in this synopsis is the ViT-ShufflenetV3 model. Effective detection is done using the ViT-assisted ShuffleNetV3 model. The outcomes obtained from the executed ViT-ShufflenetV3 model are considered for experimentation. Various conventional approaches are used for

comparing the deployed skin cancer detection model's performance. Enhanced performance is shown by the designed skin cancer detection model on experimentations.

3.1 SKIN CANCER DATASET: DETAILED DESCRIPTION

The dermoscopy images that are collected for executing the developed skin cancer detection model are gathered from the dataset. A detailed description of the gathered images is provided in Table 3.1.

Table 3.1: Description of the Gathered Dermoscopy Images.

Dataset Name	Dataset Link	Dataset Description
"HAM10000 dataset"	"https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T access date: 2023-11-18"	A total of around ten thousand dermoscopy images are available in this dataset.

The collected images are represented as DI_{di}^{coll}. Some of the sample images are given in Figure 3.2.

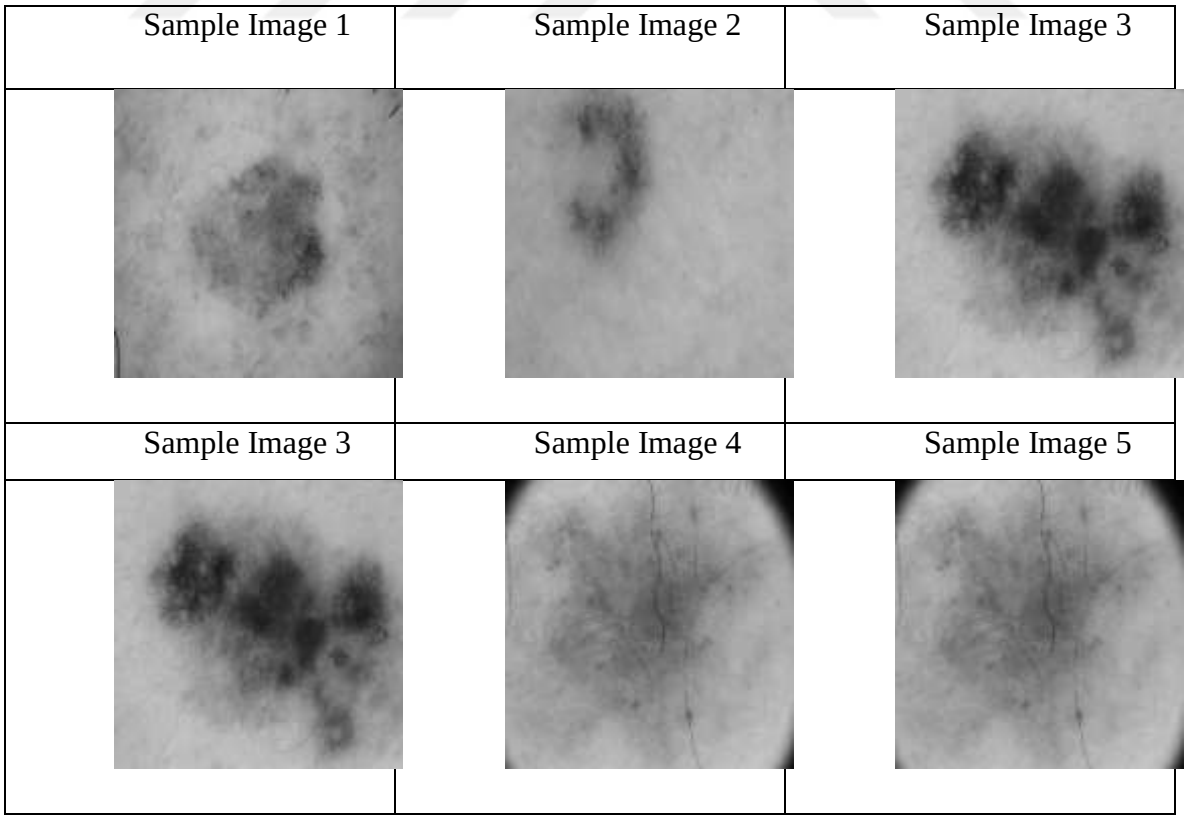


Figure 3.2: Sample Images Obtained from the Dataset.

Biomedical image segmentation can be achieved with U-Net [32], which is a sophisticated network. It has two paths: one that contracts on the left and expands on the right, creating a symmetrical U-shaped structure. The connection between every identical layer of both pathways is carried out with the aid of a few map-channels. Every channel map serves as a conduit for the communication of localization and contextual data. There are four stairs on the left-hand narrowing path in which a down-sampling operation is performed. Each step employs the "Rectified Linear Unit (ReLU)" function as the rectification function and consists of a pair of convolutional layers of size 3×3 followed by a maximum pooling layer of size 2×2. The context can be captured by training this path. The matching four stairs also make up the wide right-hand path. Here, an up-sampling operation is performed. Every stage makes use of the ReLU for rectification in addition to a pair of layers of convolution and an up-sampling layer. The feature map representations generated by the encoding sub-unit are obtained directly by the decoding component in U-Net. The process that took place on the left-hand side represents the down-sampling which is the encoding procedure, whereas on the right-hand side, the up-sampling process is conducted. The down-sampling process is also called the contraction phase. The given input image features are obtained and the encoding of the feature maps is carried out on the encoding side, while on the decoding side, these feature maps are decoded to generate the required output image. This phase is also denoted as the expansion phase. A conventional UNet is illustrated in Figure 3.3.

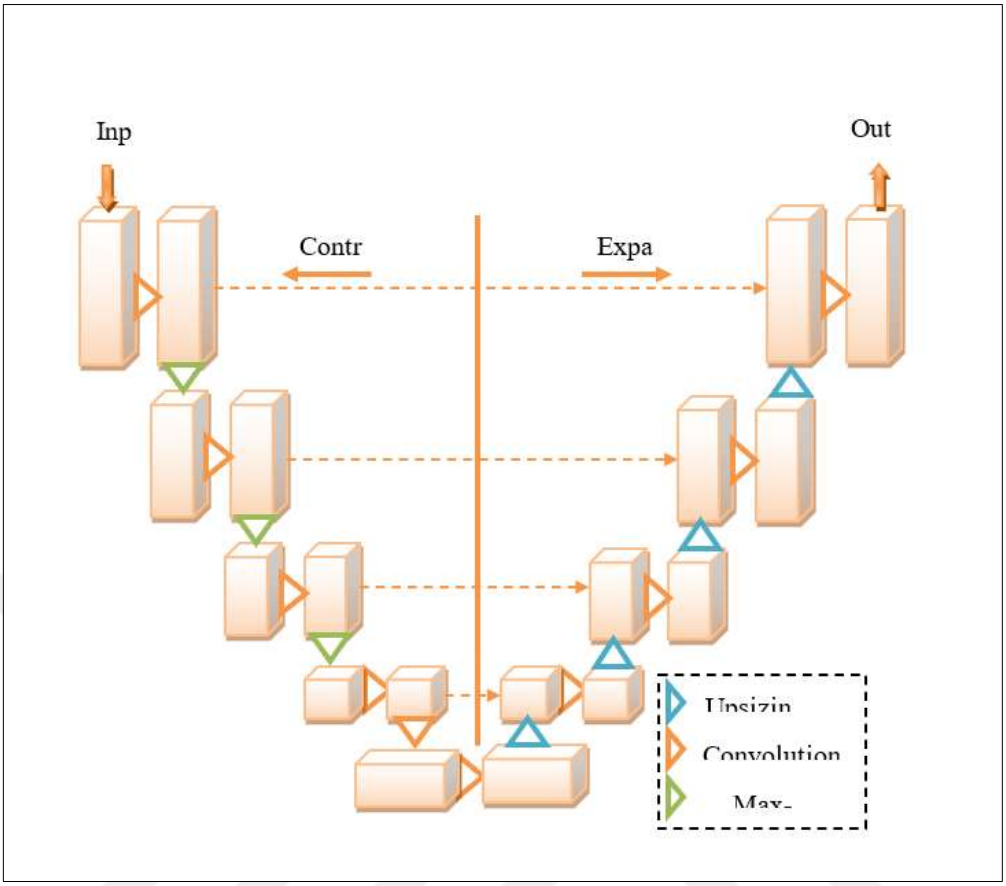


Figure 3.3: A conventional UNet is Illustrated.

UNet++ [33], begins with a backbone, or encoding sub-unit, and then moves on to a decoding sub-unit. The modified skip paths that link the two different sub-units and the application of extensive supervision set on the UNet++ is what makes the UNet++ distinct from the UNet model. The encoding and decoding sub-units' connection is altered via modified skip paths. The feature map representations generated by the encoding sub-unit are obtained directly by the decoding component in U-Net, whereas in UNet++, they get transmitted via a densely connected convolutional unit, whose number varies depending on the level of the pyramid. For instance, a densely connected convolutional unit with three layers of convolution makes up the skip pathway between two nodes. Prior to every convolutional layer, a concatenation layer that merges the prior dense unit's output from the convolutional layer with the dense unit of the lower layer that has been up-sampled takes place. In simple terms, the densely connected convolutional unit raises the encoding unit's feature maps' semantic range to the degree of feature maps that are ready to be decoded by the decoding unit. The idea behind this technique is that whenever the data obtained from

the encoding unit's feature maps and the matching decoding unit's feature maps become semantically comparable, the optimizer will have an easier time solving the optimization issue. The structure of UNet++ is given in Figure 3.4.

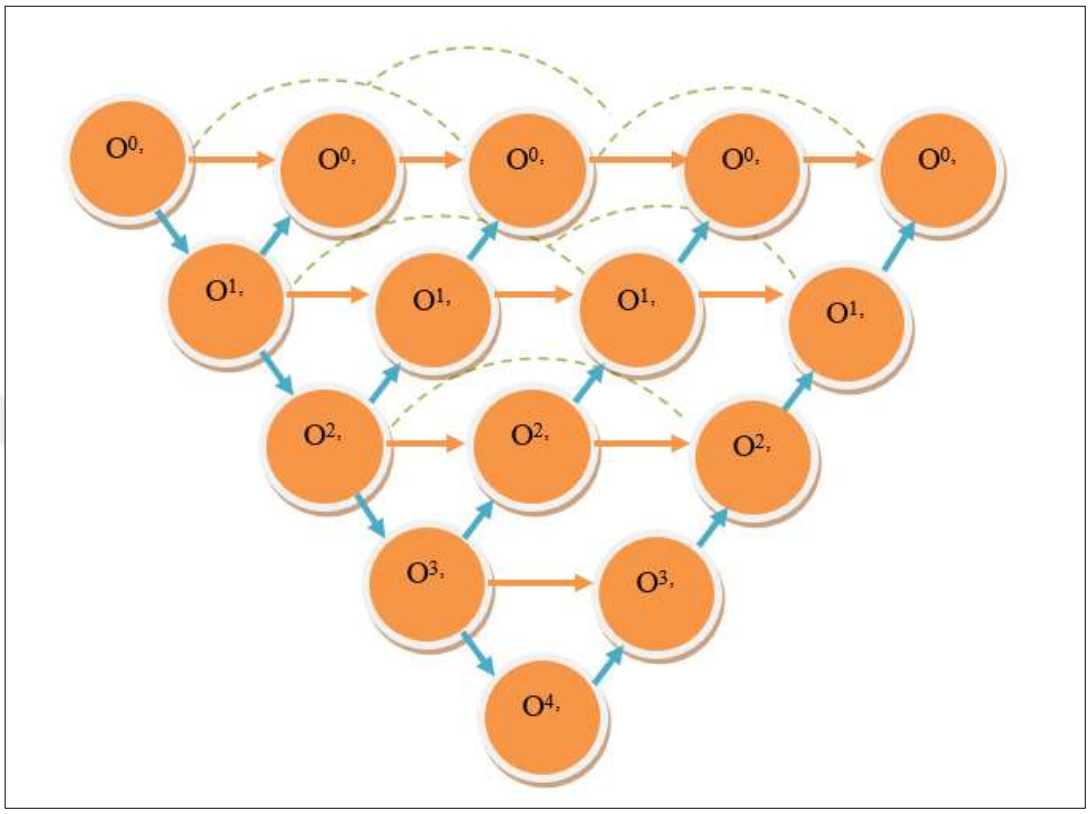


Figure 3.4: Structure of UNet++.

The various units such as $O^{0,0}$ is connected with the unit $O^{1,3}$ by means of skip connections. The output obtained from $i^{l,m}$ is denoted by $O^{l,m}$. The term m denotes the convolution layers that are present in the skip connections and l indicates the down-sampling layer's index. The obtained feature maps from $i^{l,m}$ is computed using Eq. (3.1).

$$i^{l,m} = \begin{cases} \beta\left(\left[i^{l,j}\right]_{j=0}^{m-1}, \chi\left(i^{l+1,m-1}\right)\right) & 0 < m \\ \beta\left(i^{l-1,m}\right) & 0 = m \end{cases} \tag{3.1}$$

In Eq. (3.1), the symbol $[\,]$ indicates the concatenation layer, the term $\beta(\cdot)$ denotes the convolutional function, and $\chi(\cdot)$ indicates the up-sampling layers. On $m = 0$ level, the nodes obtain a single input obtained from the encoder's prior layer. On $m = 1$ level, the nodes get a pair of inputs coming from the encoder unit, but at two different levels. When the level is

$1 < m$, the nodes obtain $1 + m$ number of inputs, such that m are the outcomes generated by the prior nodes in the exact same skip pathway and the final input is the up-sampled results from the bottom skip connection.

The skip pathway is formalized as provided in Eq. (3.2).

$$\delta\left(P, \hat{P}\right) = \frac{1}{Q} \sum_{k=1}^{Q} \left(\frac{1}{2} \cdot P_k \cdot \log \hat{P}_k + \frac{2 \cdot P_k \cdot \hat{P}_k}{P_k + \hat{P}_k} \right) \tag{3.2}$$

In Eq. (3.2), the term Q indicates the batch size, P_k represents the flattened anticipated probability of the k^{th} image, and $\hat{P}_k$ denotes the flattened ground truth.

Resolution loss usually occurs because in UNet++ as it does four down-sampling operations prior to the concatenation process. Because of this resolution loss, extensional strategies relying on an intricate network instead of a shallower one are needed to increase overall accuracy. Thus, dense blocks are included into the UNet++ structure forming the Dense-UNet++ framework.

A denser down-sampling path on the right-hand side and a denser up-sampling path on the left-hand side make up the DenseUNet++ [34] structure. In the DenseUnet++ structure, both paths are symmetrical to each other. To combine the two pathways, several bypass channels are introduced into the network. The architectural view of the DenseUNet++ is shown in Figure 4.5.

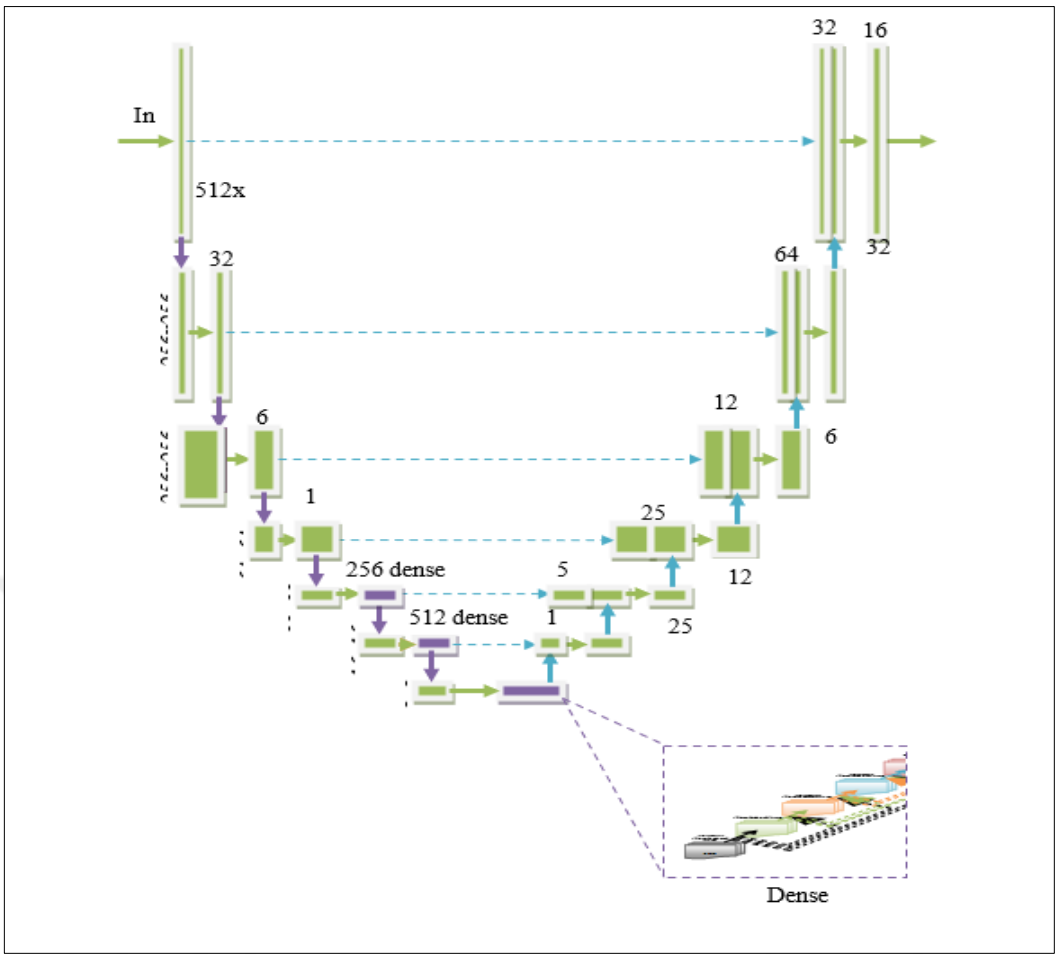


Figure 3.5: Architectural View of the DenseUNet++.

The convolutional operation in the denser down-sampling path helps to obtain the semantic contextual data on various scales. The input mages have a lower resolution because it has only four down-sampling layers which are used for gathering the local features. A lower resolution of $128 \times 128 \times 3$ is regarded as the size of the input image. In order to overcome the limiting layers in the UNet design, the dense blocks, and the transition blocks are substituted into the UNet++ structure to deepen the network's depth and replace the convolutional and pooling operations that are conventionally performed in the down-sampling path. This path is thus regarded as the denser down-sampling path. To maximize the reuse of features, every single layer in the dense blocks is connected in a feed-forward manner with all of its preceding layers. In particular, each layer's input that is considered as the dense concatenation is capable of receiving the resultant feature maps from all of the previous layers. It additionally necessitates the need for the dense blocks in every layer's feature maps to have an equal feature size. Ten dense blocks make up the suggested DenseUNet++ model

in which five units are dedicated to performing the denser down-sampling operation and five units are dedicated to performing the denser up-sampling operation. Every dense block is made up of four tightly interconnected layers with the exact same size of features. In order to accomplish the transition of layers, the transition block is provided. The denser down-sampling path consists of five levels, each consisting of "a transition block and a dense block". To recreate the higher-quality images in the denser up-sampling operation, "a dense block, up-sampling layer, and a merging operation" is used. Five layers make up the route, which recovers the entire resolution of the input and localizes the areas. On the other hand, the dense up-sampling approach closely resembles the UNet process when the dense block is used in place of the convolutional layers. A dense block contains four densely linked levels. These levels are denoted as dense convolutional layers. There are two processes of convolutions that take place in every dense convolutional layer. A dense block contains a total of eight convolutional functions in total. Every layer is linked with every previous layer in order to maximize the utilization of the retrieved characteristics and compensate for the resolution loss.

This can be illustrated by considering the feature maps from the first, second, and third layers being received by the fourth dense convolutional layer. To prevent adding more parameters to the system, the denser linkages are only used within the dense blocks. There are a number of processes for each dense convolutional layer, such as "dropout, ReLU activation function, convolutional operation, and Batch Normalization (BN)". To reduce the vanishing gradient issues, the BN procedure is initially performed on the given input image. Subsequently, the outputs undergo the process of convolution using a kernel of size1×1in order to reduce the quantity of the input feature maps. For obtaining the features, additional down-sampling using a convolutional function is performed along with the inclusion of a kernel of size3×3. To prevent overfitting issues, the dropout layers are included next. For the fusion to be effective, the transition blocks are applied between the two layers within the dense down-sampling operational path. This path also encompasses a maximum pooling operation using a kernel of size2×2, a convolutional operation, and a BN operation. To reduce the overall dimension of the output, the filter shifts two pixels at a time due to the stride of two maximum pooling operations. The dense block, transition block, up-sampling convolutional layers, and merge layers are the four fundamental modules that make up the DenseUNet++ structure.

3.2 DEVELOPED RESDENSEUNET++-BASED SEGMENTATION NETWORK

The collected dermoscopy images DI_{di}^{coll} are provided as input to the implemented ResDenseUNet++ model for the segmentation procedure. The ResDenseUNet++ is developed ResNet with DenseNet forming the "Residual Dense Network (RDN)" structure. This RDN is given to the UNet++ structure to form the ResDenseUNet++.

ResNet [35]: The ResNet model is a special type of network that makes use of hop connections to achieve residual learning. As the number of filters in the ResNet is less, it seems to be a non-complicated model when compared with the conventional VGG models. The main aim behind the development of the ResNet is to tackle the disappearing gradient issues. This goal is achieved by means of utilizing hop connections that permit the free flow of gradients without any disturbances. Some of the ResNet variants include "ResNet152, ResNet101, and ResNet50". The bypassing of the main layers of the networks is done by utilizing the hop connections in the ResNet. These residual connections help in propagating the gradients to the prior layer from the output layer without many complications. The number of hop connections used in the ResNet can be "$1, 2,$ or 3". The training process of CNN often deals with disappearing gradient issues as the preceding layer's gradients descend to zero when the model gets trained. These hop connections thus aided in the faster training of the network along with aiding in the disruption-less gradient flow. Also, these hop connections allow in expanding the layers in the network. Identity connections are added to the network to enhance the performance. This assisted in faster and easier training by making the gradient's back-propagation an effortless process. The two main units of the ResNet are listed as follows:

a. Convolution unit
b. Identity unit

Convolutional Unit: Distinctly sized outcomes and inputs are present in this unit of the ResNet model. Here, the linear projection is carried out to eliminate the deviation in sizes between the output and input. The structure of the convolutional unit of the convolutional unit and the identity units are the same. But, an extra convolutional layer of 2D is added in the hop connection of the convolutional unit of the ResNet. This hop connection is the one where the size of the input is equal to that of the output and is provided to the main path. A filter of size 1×1 having a stride on the basis of the size of the required output is present in the

2D convolutional layer. This aids in resolving the vanishing gradient issues. This makes sure to obtain enough knowledge from the identity unit of the ResNet.

Identity Unit: An ideal identity unit of the ResNet structure consists of the following three layers.

a. A filter of size 1×1 having a padding of $(1,1)$ is considered as the primary unit in the convolution layer of two dimensions. In this layer, the ReLU activation is performed, and the channels get normalized using the "Batch Normalization (BN)" function.
b. On the second unit, the same component is provided as the prior unit, but the size of the filter is changed as $s\times s$.
c. A layer similar to the initial unit but having no ReLU activation operation forms the third unit of the ResNet.

At last, a ReLU activation is performed before which the concatenation of the input and the hop connections take place.

Initially, the ResNet is connected with the DenseNet structure forming the "Residual Dense Network (RDN) [35]," structure. The major contributions of the RDN are as follows.

Backbone Network: The backbone network for the RDB is created by introducing a residual framework into the inner dense block of the structure. The model that is used as the dense block is the DenseNet-121 network, which was pre-trained on ImageNet. The medical image serves as its input, and the feature maps that are taken out of the image serve as its output.

"Feature Output Unit (FOU) [36],": The pictorial illustration of the FOU is given in Figure 6. It has three layers, which include a "global mean pooling layer, a 1D convolutional layer, and a 2D convolutional layer having filters of size". A 64 -length feature vector is produced following the processing of this unit and the feature mappings are extracted in the backbone network. The given input medical image's robust feature vector is what is extracted from these units as output. This process is illustrated in the below provided figure 3.6.

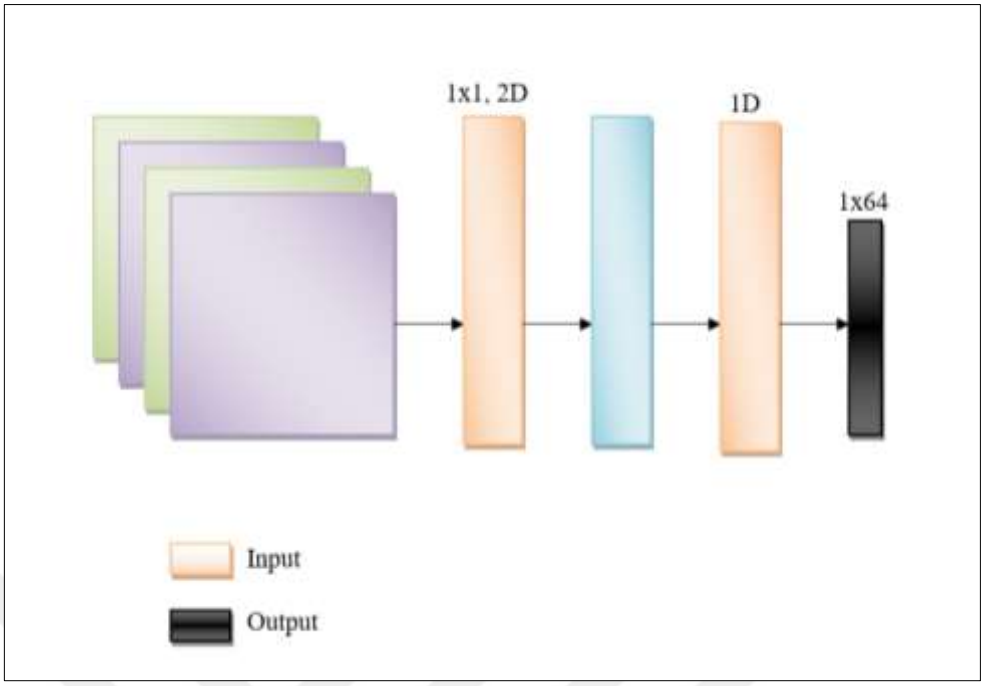


Figure 3.6: Pictorial Illustration of the FOU of the RDN.

The feature maps of distinct scales are extracted by means of the DenseNet structure that is being utilized in the RDN structure. This DenseNet model is highly potent in feature extraction capability. By connecting two dense blocks via skip connection, one may completely extract the deep semantic connections between the provided input medical images, and these high-level features can exhibit better robustness when compared to low-level features. The system's resilience is directly correlated with how well the model extracts features from the medical images. Therefore, in the FOU layer, 1D convolution is used to produce a resilient feature vector of size 64 , global mean pooling is employed to minimize dimensionality and 1×1 convolutional operation is performed to decrease the feature mapping count. To produce a robust feature vector, use a mean binarization operation on the robust features that are obtained from the RDN model for generating the hash function of these feature vectors which is determined as given in Eq. (4).

$$S(n)=\begin{cases}0 & \chi \le R \\ 1 & \chi > R\end{cases}, \chi = \frac{1}{64}\sum_{n=0}^{63} R(n) \tag{3.3}$$

In Eq. (4), the term $S(n)$ denotes the hash vector and $R(n)$ indicates the feature vector. A total of four units are present in the RDN model. These units are “Up-sampling Network (UPNet)”, “Shallow Network (SNet)” for feature extraction, “Dense Feature Fusion (DFF)”,

and “Residual Dense Block (RDB)”. The input given to the RDN is denoted by C_{GH} and the obtained output is represented by means of C_{DH}. The deep features that are extracted by means of the SNet are obtained with the aid of a pair of convolutional layers. From the given input, the E_{-1} features are extracted with the aid of the first convolutional layer. The extracted features from the first convolutional layer are computed using Eq. (3.3).

$$E_{-1} = B_{DEF1}(C_{GH}) \tag{3.3}$$

In Eq. (3.3), the term $B_{DEF1}(\cdot)$ indicates the convolutional operation. Then with the utilization, the deeper features E_{-1} are further extracted and the associated Global Residual Learning (GRL) is performed. The deep features that are extracted from GRL are given by Eq. (3.4).

$$E_0 = B_{DEF2}(E_{-1}) \tag{3.4}$$

In Eq. (3.4), the term $B_{DEF2}(\cdot)$ indicates the convolutional operation that is performed on the second convolutional layer of the SNet. The extracted features from the second convolutional layer E_0 are provided as input to the RDB. Let us consider a total of I RDBs are present in the RDN model. Therefore, the features E_a are extracted as output from the a^{th} RDB unit.

$$E_a = B_{HIJ,a}(E_{a-1}) \tag{3.5}$$

$$E_a = B_{HIJ,a}(B_{HIJ,a-1}(\cdots(B_{HIJ,1}(E_0))\cdots)) \tag{3.6}$$

In Eq. (3.5) and Eq. (8), the term $B_{HIJ,a}$ represents the operation performed on the RDB that is present in a^{th} unit. A series of operations are denoted by $B_{HIJ,a}$, in which “Rectified Linear Unit (ReLU)” and the convolution operation are performed. The term a^{th} is used to represent the local features that are generated by a^{th} RDB, which completely utilizes each of the block’s convolutional layers.

Then the operations such as "Dense Feature Fusion (DFF)", which combines GRL and “Global Feature Fusion (GFF)" following the extraction of the complex characteristics using a set of RDBs is performed. By fully exploiting the characteristics from all prior levels, the feature maps are generated by the DFF as given in Eq.(3.7).

$$E_{IE} = B_{IAA}(E_{-1}, E_0, E_1, ..., E_I) \tag{3.7}$$

In Eq. (3.7), the term E_{IE} represents the resultant feature-maps from DFF that are obtained using an integrated function represented by B_{IAA} on the DFF unit of the RDN. The various numbers of "Up-sampling Networks (UPNets)" are arranged over one another in the HR region once the global and local features are extracted from the LR region. A single convolutional layer following the ESPCN [37] is present in the UPNet which is given by Eq. (3.8).

$$C_{DH} = B_{HIK}(C_{GH}) \tag{3.8}$$

In Eq. (3.8), the term B_{HIK} indicates the function that is used to retrieve the outcome from the RDN."Local Residual Learning (LRL), Densely Connected Layers (DCL), and Local feature fusion (LFF) make up the RDB, which results in the mechanism of "Contiguous Memory (CM)". Every single layer in the present RDB receives the prior RDB's state in order to implement the CM technique. Let E_{a-1} denotes the DRB's input and E_a indicates the DRB's output, respectively, with L_0 feature maps for each. The final result generated by the b^{th} convolutional layer of the a^{th} RDB is provided as follows in Eq. (3.9).

$$E_{a,b} = \alpha\left[M_{a,b}\left(E_{a-1}, E_{a,1},..., E_{a,b-1}\right)\right] \tag{3.9}$$

The term α in Eq. (3.9) stands for the activation function used, i.e., ReLU. The relative weights within the b^{th} convolutional layer are represented by $M_{a,b}$. The bias term is neglected while considering the simplicity purpose. Let us consider that the growth rate L feature maps make up $E_{a,b}$. The concatenation of the feature-maps generated by $(a-1)^{th}$ RDB is represented as $\left(E_{a-1}, E_{a,1},..., E_{a,b-1}\right)$, convolution layers $1, 1,..., (b-1)$ available in the RDB present at the a^{th} position generates feature maps as $L_0 + (b-1)\times L$. Every layer's results and those of the previous RDB are directly connected with every one of the layers that come after it, maintaining feed-forward functionality while also extracting local dense features. The states from the previous RDB and all of the convolutional layers in the present RDB are then automatically fused via local feature fusion. As previously examined, the $(a-1)^{th}$ RDB's feature-maps are concatenated directly to the a^{th} RDB; hence, reducing the feature count is a crucial factor. In contrast, a 1×1 convolution layer is incorporated into the unit, for regulating

the output data in an adaptive manner, drawing inspiration from MemNet [38]. This process is referred to as LFF which is expressed as provided in Eq. (3.10).

$$E_{a,GA} = B^{a}_{GEE}\left(E_{a-1}, E_{a,1},..., E_{a,b},..., E_{a,N}\right) \tag{3.10}$$

In Eq. (3.10), the term B^{a}_{GEE} represents the a^{th} RDB's 1×1 convolutional layer's operation. It is observed that it would be difficult in training an extremely deep dense network without LFF when the growth rate L increases. Since an RDB contains many convolutional layers, LRL is added to enhance the information flow even more. The a^{th} RDB's ultimate output can be acquired as mentioned in Eq. (3.11).

$$E_{a} = E_{a=1} + E_{a,GE} \tag{3.11}$$

It has to be mentioned that LRL can enhance the representation of the network even further, leading to improved performance. As LRL and dense connectivity are utilized, this unit is referred to as RDB.

The pictorial representation of RDN is given in Figure 3.7.

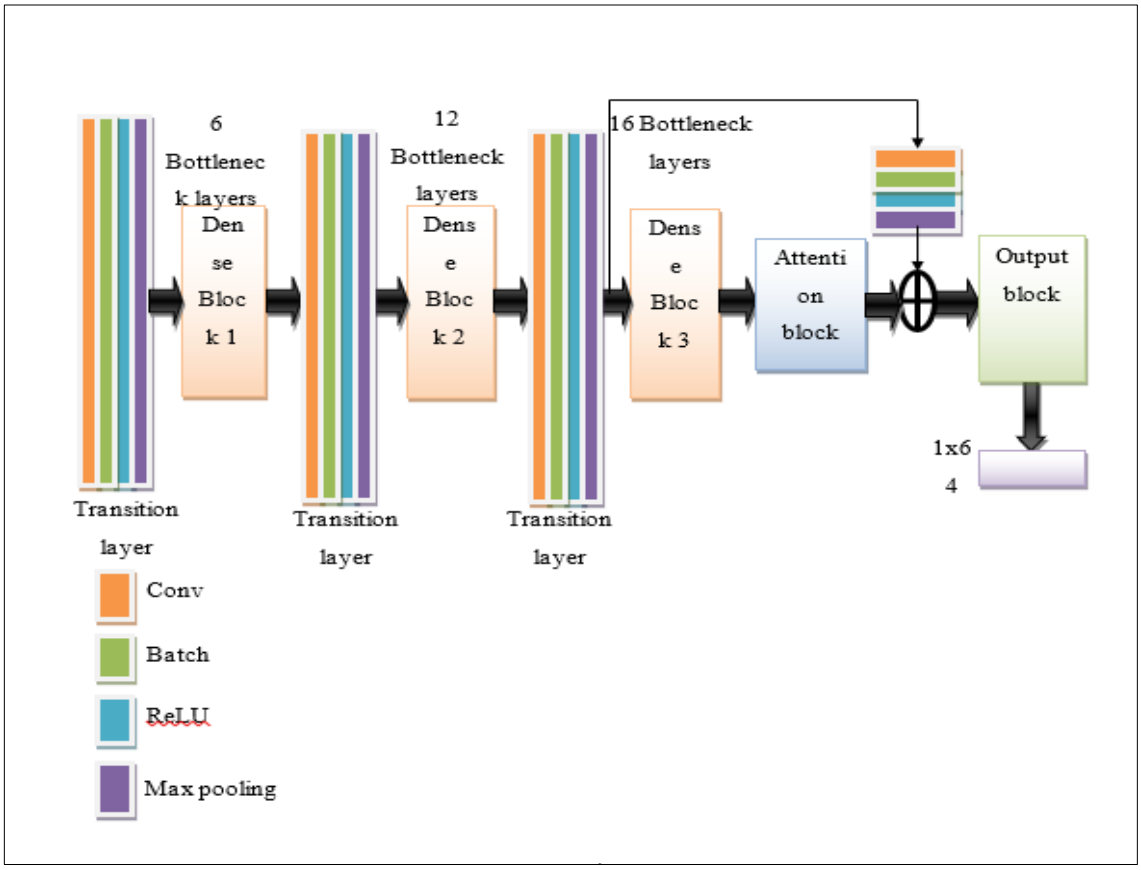


Figure 3.7: Pictorial Representation of RDN.

Denser fusion of features: The DFF is suggested to leverage the complex characteristics globally following the extraction of the LDF using a collection of RDBs. GFF and "Global Residual Learning (GRL)" make up the DFF structure of the RDN. By combining features from all RDBs, GFF is suggested as a means of extracting the global feature FGF as given in Eq. (3.12).

$$E_{LE} = B_{LEE}(E_1,..., E_I) \tag{3.12}$$

In Eq. (3.12), the term $(E_1,..., E_I)$ denotes the concatenated feature maps generated by the RDB. A composite function comprising a 1×1 and 3×3 convolutional operation is called HGFF. A variety of characteristics with varying degrees are adaptively fused using the 1×1 layer of convolution. It has been shown to be effective in further obtaining features for GRL by

adding subsequent 3×3 layers of convolution. After that, the feature-maps are obtained using GRL, and upscaling is done which is shown in Eq. (3.13).

$$E_{IE} = E_{-1} + E_{LE} \tag{3.13}$$

The term E_{-1} in Eq. (3.13) indicates the low-level feature-maps. The prior layer that exists before the GFF is fully utilized by the developed RDB. Multiple levels of "Local Dense Features (LDF)" are generated by RDBs and then subsequently combined to form FGF through adaptive means. Following GRL, a dense feature from the FDF is obtained. Long-term interconnections are limited by the incomplete usage of local feature data. Furthermore, MemNet gathers information in the HR domain, thus boosting the complexity of computation. Here, the LR region is used to obtain the local and global characteristics.

Model execution: All convolutional layers in the residual unit of the RDN have a size of 3×3 , with the exception of the one used for local and global feature fusion, which has a kernel of size 1×1. To maintain a stable size for all the layers of convolution with a kernel size of 3×3, zero padding is done on both sides of the input. Filters $L_0 = 64$ are used in the deep feature extraction, local, and GFF layers. Filters L are used in the other layers of each RDB, and ReLU comes next. The up scaling of the lower resolution characteristics to fine ones within UPNet using ESPCNN is then carried out. Since we are producing color HR images, the last convolutional layer has three output channels. But the network is also capable of processing grey-scale images as well.

The UNet++ structure is combined along with the RDN forming the ResDenseUNet++ structure which is illustrated in Figure 8. The segmentation outcomes are obtained from the ResUNet++ model which is denoted as SI_{si}^{RDU} . From the segmented images, the text pattern images TP_{tp}^{images} and the Gabor filter images GF_{gf}^{images} are obtained as the output. These two images are further given to the detection model of skin cancer.

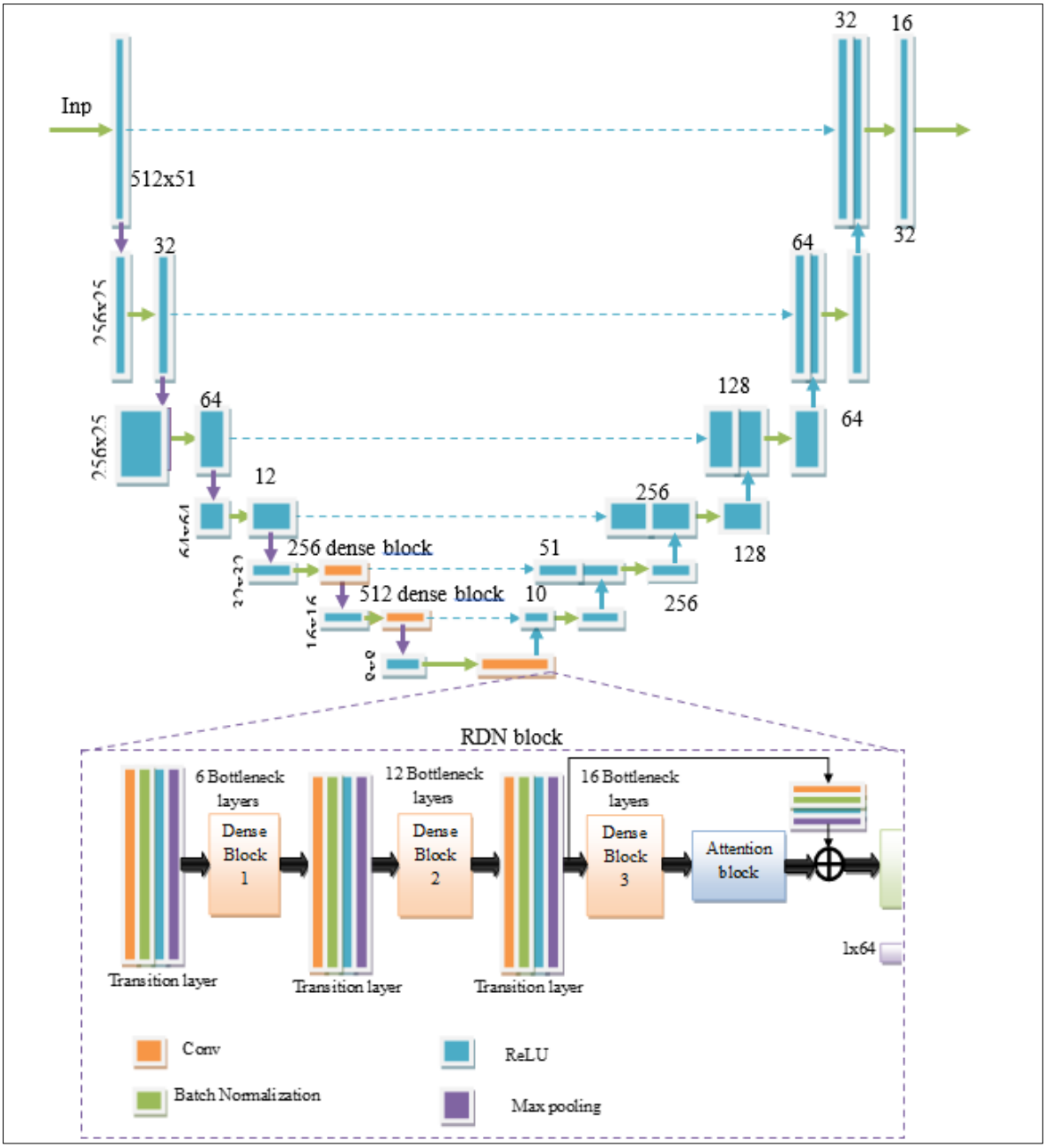


Figure 3.8: Diagrammatic Representation of ResDensUNet++.

CNN models are utilized in various image-processing tasks. However, the computational complexity of the network increases as the depth of the layer's increases. Also, the network becomes more weighted. A weightless CNN model with efficient computation capability developed to overcome these limitations is called the ShuffleNet [39]. ShuffleNet is computationally faster and has the highest accuracy rate in various image processing operations. The unique nature of the ShuffleNet is its channel shuffle operation. The ShuffleNet separates the channel groups by rearranging the input channel's data and making

the output from every group closely imitate the input. Along with DSC operation, the ShuffleNet model is made user-friendly. Both the channel-wise as well as the spatial characteristics are captured by this DSC without much computation processes. The two main functions performed by the ShuffleNet model are channel shuffle and point-wise group convolution [40].

Group Convolution: The process of assigning the network to two "Graphical Processing Units (GPU)" for enhancing the effectiveness of the model is called group convolution. In traditional convolution, the convolutional operation makes use of the dense connection of channels for executing a convolutional operation on every input characteristic's channels. The height and width of the kernel in a traditional convolutional layer is denoted by I , and the number of channels in the input is represented as A . When A becomes J , then the output channel count also becomes J . The parameters required by the traditional convolution T_W are given by Eq. (3.14).

$$T_W = A \times A \times U \times V \tag{3.14}$$

In the case of group convolution, the input characteristic's channels are partitioned into X number of groups so that the kernel count becomes $\frac{U}{X}$. The outputs obtained from all the groups X are summed up to generate a huge output feature with J number of channels. The parameters required by the group convolution T_Y are given by Eq. (3.15).

$$T_Y = A \times A \times \frac{U}{X} \times V \tag{3.15}$$

The above illustration proves that the parameter requirement of the group convolution is lesser than the traditional convolution.

Channel Shuffle: Even though the group convolution decreases the parameter requirement of the ShuffleNet, it also has the drawback of reducing the possibility of sharing data between distinct groups in the network. This drawback is rectified by the introduction of the channel shuffle operation in the ShuffleNet. The channels of the output characteristics are shuffled in the ShuffleNet structure by the channel shuffle operation so that the data gets passed through various groups in the ShuffleNet without much computational cost. The diagrammatic representation of channel shuffle operation is depicted in Figure 3.9.

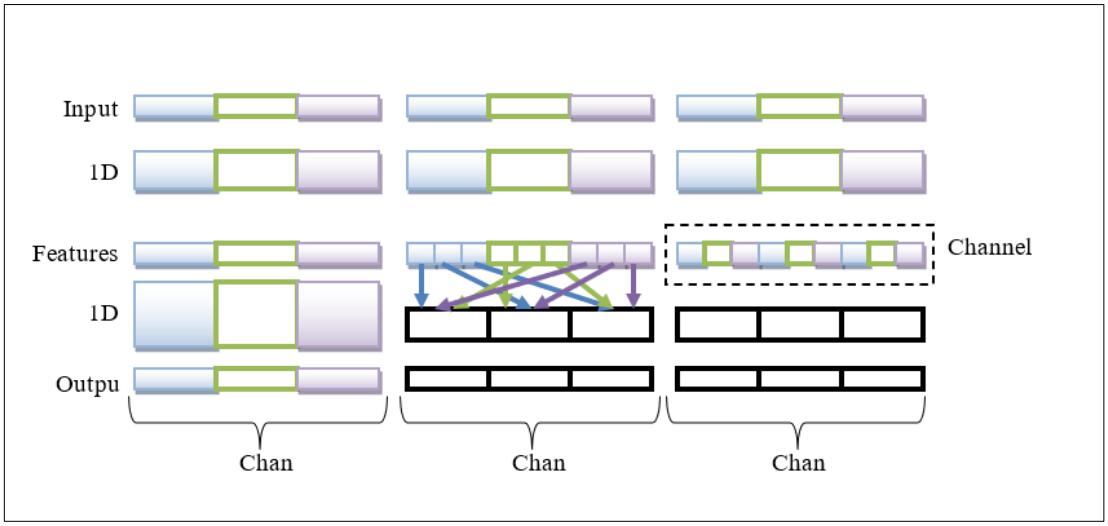


Figure 3.9: Diagrammatic Representation of Channel Shuffle Operation.

The building components that comprise the contemporary CNN are typically repeating and have an identical structure. "Xception and ResNeXt" are two cutting-edge systems that use group convolutions or efficient "Depthwise Separable Convolutions (DSC)" in their fundamental components to achieve a superior balance between computational cost and representative capacity. It is important to note that both of these designs fail to properly account for 1×1 convolutions, which are also known as pointwise convolutions, and result in a significant amount of complexity. Expensive pointwise convolutions in small networks lead to a restricted number of channels to satisfy the degree of complexity restriction, which may seriously impair accuracy. Applying channel sparse connections, such as DSC on 1×1 layers is a simple way to solve this problem. DSC dramatically lowers computation costs by guaranteeing that each convolution only affects the relevant source channel group alone. One of the drawbacks that exist in the DSC though is its occurrence when several DSCs are stacked together when the outputs from a given channel are only obtained from just a few percent of input channels. It is evident that a group's outputs are only related to the group's inputs. This characteristic reduces representation by obstructing the exchange of data between channel groups. Enabling DSC to acquire input data from distinct groups will result in a complete relationship between both the input and output channels. To be more precise, split the channels in every group into multiple subgroups for the feature map produced by the prior group layer, and then supply every group in the subsequent layer with a distinct sub-group. A channel shuffle procedure can be used to accomplish this in an attractive and

effective manner. Consider a layer of convolution with g groups and $c \times d$ channels in its output. To transform the dimension of the output channel back into (c, d), transpose it, and then flatten it again to use as the input for the following layer. Keep in mind that regardless of whether the two convolutions contain distinct sets of groups, the operation still occurs. Channel shuffling can also be inserted into network topologies for automated training because it is also distinguishable. Multiple group layers of convolution can be used to create more potent structures as a result of the channel shuffle technique.

An effective model with DSC and channel shuffling is developed in the following section. Utilizing the channel shuffle function, a new ShuffleNet [41], unit is created especially for tiny networks. The starting point of the ShuffleNet model is the bottleneck unit which serves as its principal design component. This block is the residual unit. A computationally efficient 3×3 DSC is carried out on the feature maps of the bottleneck layer present in the residual channel of the 3×3 layer. Under identical conditions, this structure is less complex than "ResNet (bottleneck design) and ResNeXt". The quantity of convolutional groupings is also higher with the least complications. Stated differently, ShuffleNet can employ broader feature maps if it has limited computational expenses. It is found that for small networks since they typically lack the channels, it is needed to handle the information in an effective manner. Furthermore, ShuffleNet only uses bottleneck feature maps for DSC to function. DSC typically has relatively low computational complexity, but it is challenging to execute on low-power portable devices in an efficient manner. This could be because it has a lower computation and less access to memory than other dense techniques. To minimize costs, DSC is employed only on the bottleneck in ShuffleNet units. The architectural view of ShuffleNet is shown in Figure 3.10

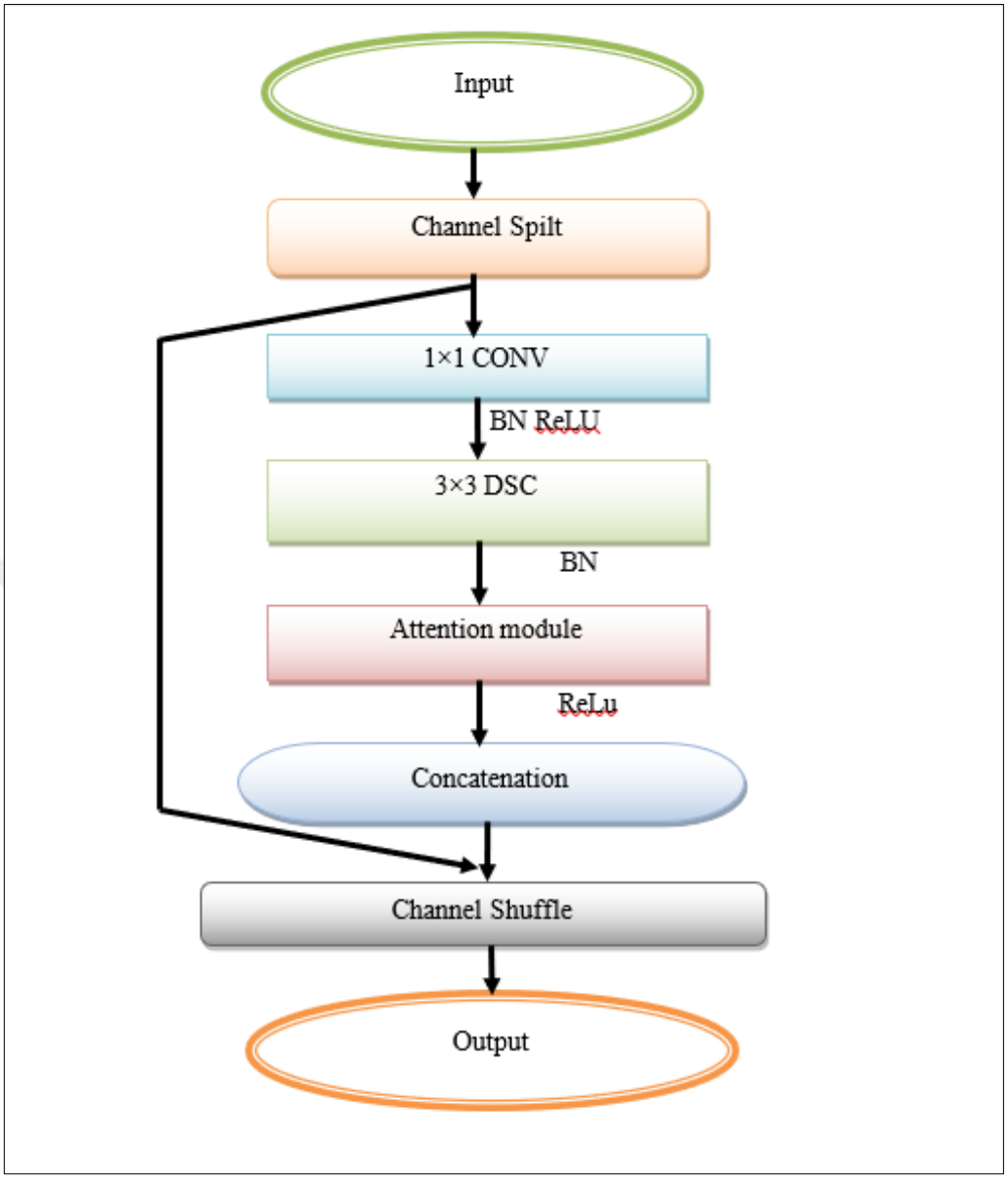


Figure 3.10: Architectural View of ShuffleNet.

The diagrammatic representation of the ShuffleNetV3 model is given in Figure 3.11.

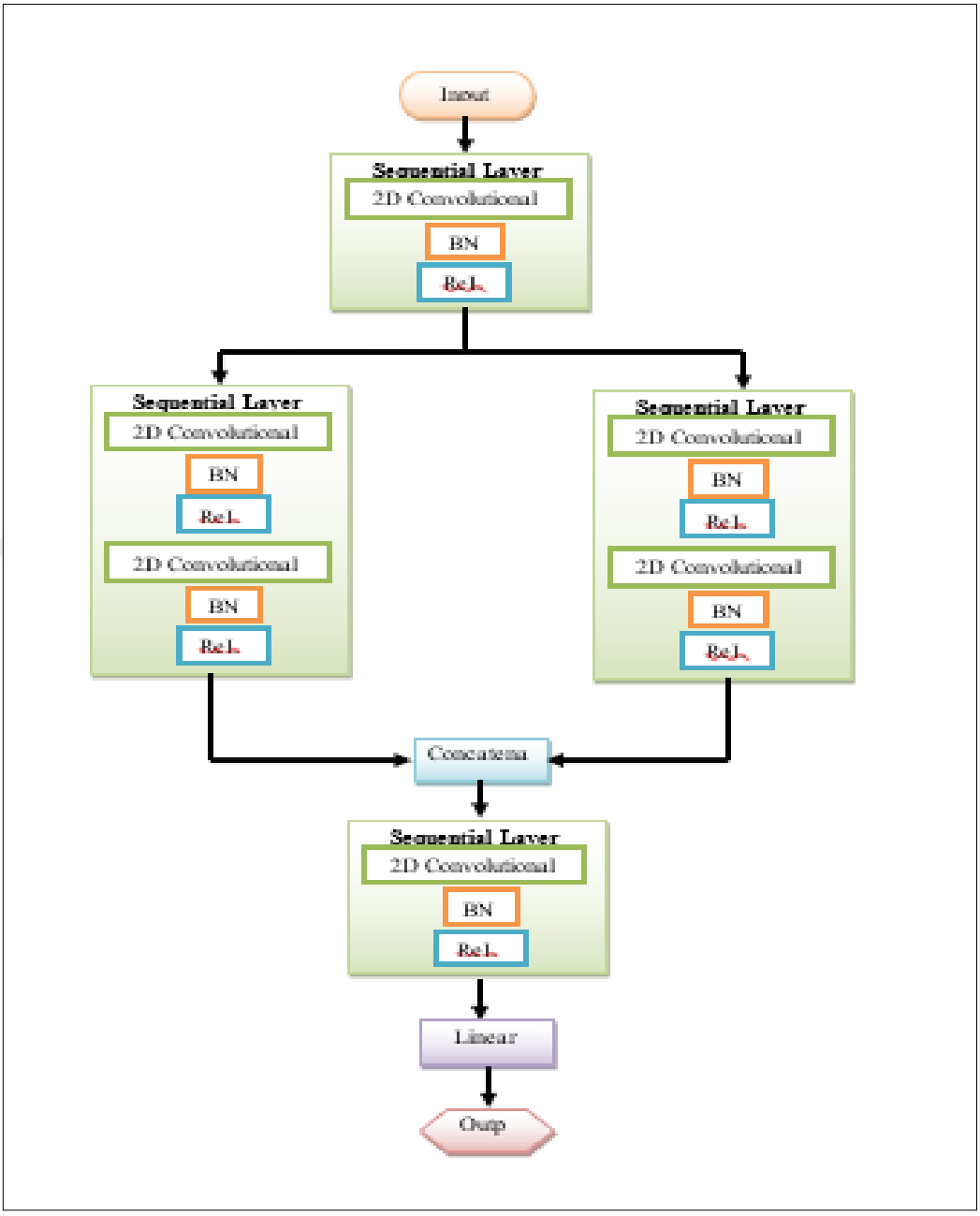


Figure 3.11: Diagrammatic Representation of ShuffleNetV3 Model.

The ShuffleNetV3 is a modified version of the ShuffleNet model. The ShuffleNetV3 is made up of several branched structures. The conventional structure of a ShuffleNetV3 model is as follows. Each branch consists of a sequential layer. In the first branch, a sequential layer consisting of a 2D convolutional layer, a “Batch Normalization (BN)”, and ReLU are present. In the second branch, a sequential layer consisting of a 2D convolutional layer followed by BN and ReLU is present. After the ReLU function of the second branch, another 2D convolutional layer is present. Again, a BN layer and a ReLU layer follow. Similar

components and layers are present in the third branch of the ShufflNetV3 model which acts parallel to the model. The output from these parallel branches is concatenated. After that another branch with a sequential layer consisting of a 2D convolutional layer, a BN, and ReLU is present. Finally, linear activation is carried out to generate the required output.

3.3 VISION TRANSFORMER: DESCRIPTION

In order to examine limited datasets in terms of classification tasks, the ViT [42], model, a variation in the conventional transformer model that is dedicated specifically to image classification tasks is introduced. ViT modifies the design of the transformer only slightly in order to make it effective in handling two-dimensional images. Transformers represent a model which works on the basis of the self-attention mechanism. Such type of transformer that works with self-attention mechanism is already considered as a conventional system in "object identification, image categorization as well as Natural Language Processing (NLP)" tasks. Transformers are widely used in NLP due to their ability to efficiently apply learning to subsequent duties. The very first transformer framework is composed of an "encoding unit and a decoding unit". This "encoding-decoding" structure is the basis for the structure of transformers that are used in image classification tasks. The encoder portion of the transformer is the only component that is used in the case of image categorization task. In order to provide input to the transformer framework's encoding component, positional encoding is applied to an image patch embedding to maintain the patch order, and the resulting "positional encoded patch embeddings" are subsequently sent to the encoding unit of the transformer. The encoding unit is made up of "feed-forward levels, Multi-Head Self-Attention (MHSA), layer normalization, and then an additional layer of normalization that is subsequently followed by a feed-forward level". Using query, key, and value matrices in the self-attention layer, attention scores are calculated to determine the relationship between every patch. To obtain a better representation, numerous self-attention ratings are calculated, acting as a multi-head. The resultants from these heads are then combined into a single vector, into which an input vector is added via a skip link, and then the normalization is carried out. The result from the aforementioned normalization layer is subsequently passed to the feed-forward level, where normalization is once again performed and the output from the previous layer's normalization is incorporated once more with the aid of bypass connections. The interaction between the representations of various levels is made possible

by these skip links. In an image classification-based transformer, different encoding units can be piled to collect the essential information. The transformer's encoding unit's final output is supplied to the classifier in order to perform classification. The output of the transformer represents the required image output.

The attention mechanism is a transformer network's key component. Connections in a series of inputs as well as outputs can be described using a system of attention regardless of how far apart, they are. "Self-attention" is a type of attention mechanism that computes the data of an entire input series by correlating multiple sequence places, and it has evolved into an essential component of persuasive sequence representation and transition approaches for a range of activities. Self-attention has been effectively used for many different activities, such as "abstract generalization, textual entailment, and reading comprehension". A mapping between a vector query, key, value, and output is the basic function of an attention mechanism. The result of the attention method is a weighted average of values, with each value's weight determined by the query's compatibility function with its matching key. "Scaled dot-product attention and MHSA" are the two types of self-attention mechanisms used in the case of a transformer. With MHSA, every attention mechanism functions as a head, and every head gain unique knowledge, improving the encoding model's representational capacity. In order to perform this multi-headed attention calculation, query, key, and value are split into Z vectors prior to implementing self-attention. Following that, each divided vector undergoes the self-attention process independently. Each stage of self-awareness is treated as a head. Each head generates a resultant vector that is combined into a single vector prior to it reaching the last linear layer. The architecture of ViT is depicted in Figure 3.12.

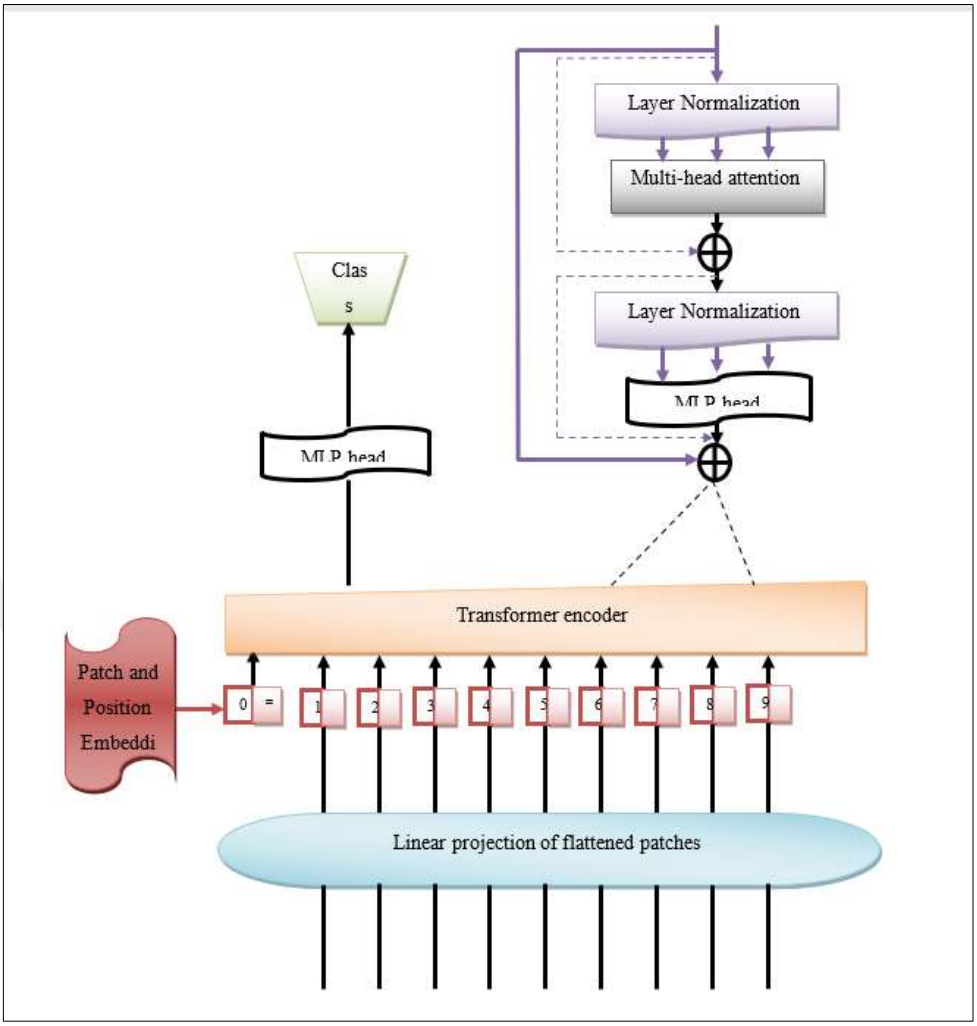


Figure 3.12: Architecture of ViT.

The operations are performed on the encoder unit of the ViT.

$$o = r \times v_s \tag{3.16}$$

$$p = r \times v_t \tag{3.17}$$

$$q = r \times v_u \tag{3.18}$$

In Eq. (18), Eq. (19), and Eq. (20), the term v_s, v_t, and v_u indicates the parameters that are used to project the patch features. The patch embedding is determined as follows.

$$o \times p^w \tag{3.19}$$

$$\frac{op^w}{\sqrt{x_t}} \tag{3.20}$$

$$y\left(\frac{op^{w}}{\sqrt{x_t}}\right) \tag{3.21}$$

$$z(o,p,q)y\left(\frac{op^{w}}{\sqrt{x_t}}\right)q \tag{3.22}$$

The SoftMax function performed on the ViT is given by Eq. (3.21). The attention function that is carried out on the transformer unit is shown in Eq. (3.22).

3.4 IMPLEMENTED VIT-SHUFFLENETV3-AIDED SKIN CANCER DETECTION

The text pattern images TP_{tp}^{images} and the Gabor filter images GF_{gf}^{images} are provided as the input to the implemented detection model. The ViT-ShuffleNet model is developed by combining the ViT into the internal layers of the ShuffleNetV3 structure. The diagrammatic illustration of the ViT-ShuffleNetV3-based skin cancer detection model is given in Figure 12. From the implemented ViT-ShuffleNetV3-based skin cancer detection, the detected skin cancer outcome will be obtained.

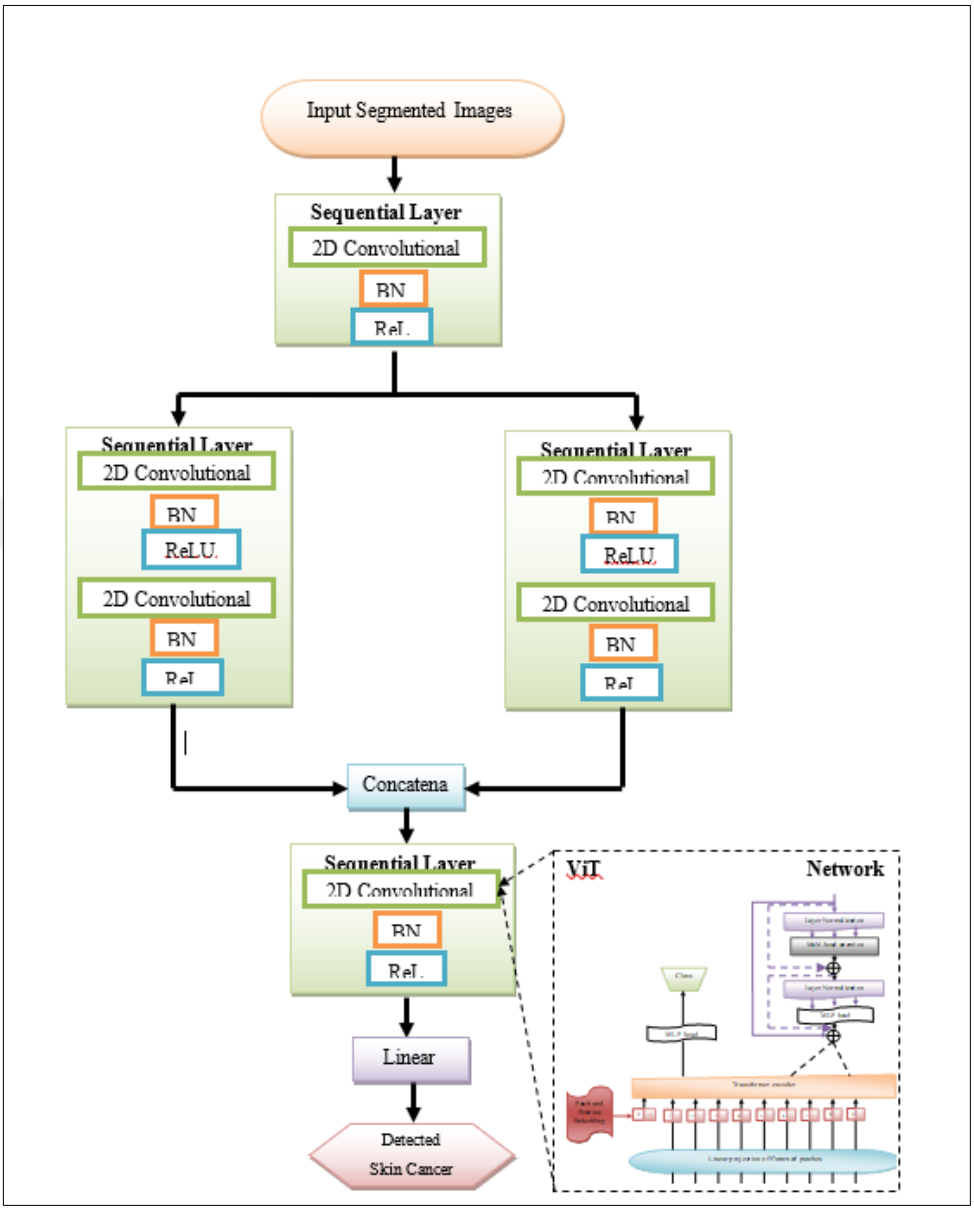


Figure 3.13: Diagrammatic Illustration of ViT-ShuffleNetV3-based Skin Cancer Detection Model.

Results and Discussion: The implemented skin cancer detection model was executed in the Python platform. The executed ResDenseUNet++-based segmentation model was compared with existing segmentation models such as "UNet [43], UNet3+ [44], Residual UNet (ResUNet) [45], and Transformer-based UNet (TransUNet) [46]. Then the executed skin cancer detection model based on the ViT-ShuffleNetV3 model was contrasted against conventional detection models such as "CNN [47], Residual Attention Network (RAN) [48], ResNet [49], and ShuffleNetV3".

3.5 VALIDATION INDICES

Various metrics are used for validating the performance of the implemented segmentation and detection models that are given as follows.

(a) The "Negative Predictive Value (NPV)" Ag is computed using Eq. (3.23).

$$Ag = \frac{\mathrm{Wc}}{Wc + Eg} \tag{3.23}$$

In Eq. (3.23), the term Wc represents the true negatives and Eg indicates the false negatives.

(b) The "False Negative Rate (FNR)" El is determined using Eq. (3.24).

$$El = \frac{\mathrm{Eg}}{Wq + Eg} \tag{3.24}$$

In Eq. (26), the term Wq denotes the true positives.

(c) The "accuracy" Ax is determined with the aid of Eq. (3.25).

$$\mathrm{Ax} = \frac{Wc + Wq}{Wc + Eb + Wq + Eg} \tag{3.25}$$

In eq. (27), the term Eb represents the false positives.

(d) The "Matthews Correlation Coefficient (MCC)" Mo is computed using Eq. (3.26).

$$\mathrm{Mo} = \frac{\mathrm{Wc} * \mathrm{Wq} - \mathrm{Eb} * \mathrm{Eg}}{\sqrt{(Wq + Eb)(Wq + Eg)(Wc + Eb)(Wc + Eg)}} \tag{3.26}$$

(e) The "dice coefficient" Qt is calculated using Eq. (3.27).

$$Qt = \frac{2 * Wq}{(Wq + Eb) + (Wq + Eg)} \tag{3.27}$$

(f) The "False Positive Rate (FPR)" Wr is computed using Eq. (3.28).

$$Wr = Wc \frac{\mathrm{Eb}}{TN + Eb} \tag{3.28}$$

(g) The value of "precision" Wj is determined with the aid of Eq. (3.29).

$$Wj = \frac{Wq}{Wq + Eb} \tag{3.29}$$

(h) The "Jaccard" Wn is calculated using Eq. (3.30).

$$Wn = \frac{Wq}{Wq + Eb + Eg} \tag{3.30}$$

(i) The "recall" Th value is determined using Eq. (3.31).

$$Th = \frac{\mathrm{Wq}}{Wq + Eg} \tag{3.31}$$

(j) The "False Discover Rate (FDR)" Sj is determined using Eq. (3.32).

$$Sj = \frac{\mathrm{Eb}}{Wq + Eb} \tag{3.32}$$

(k) The "specificity" Rk is computed with the aid of Eq. (3.33).

$$Rk = \frac{\mathrm{Wc}}{Wc + Eb} \tag{3.33}$$

(l) The "F1-score" Un is computed using Eq. (3.34).

$$\mathrm{Un} = \frac{2 * \mathrm{Wq}}{2 * (Wq + Eb + Eg)} \tag{3.34}$$

3.6 EXPERIMENTAL OUTCOMES

The segmentation outcomes from the executed ResDenseUnet++ model are shown in Figure 13.

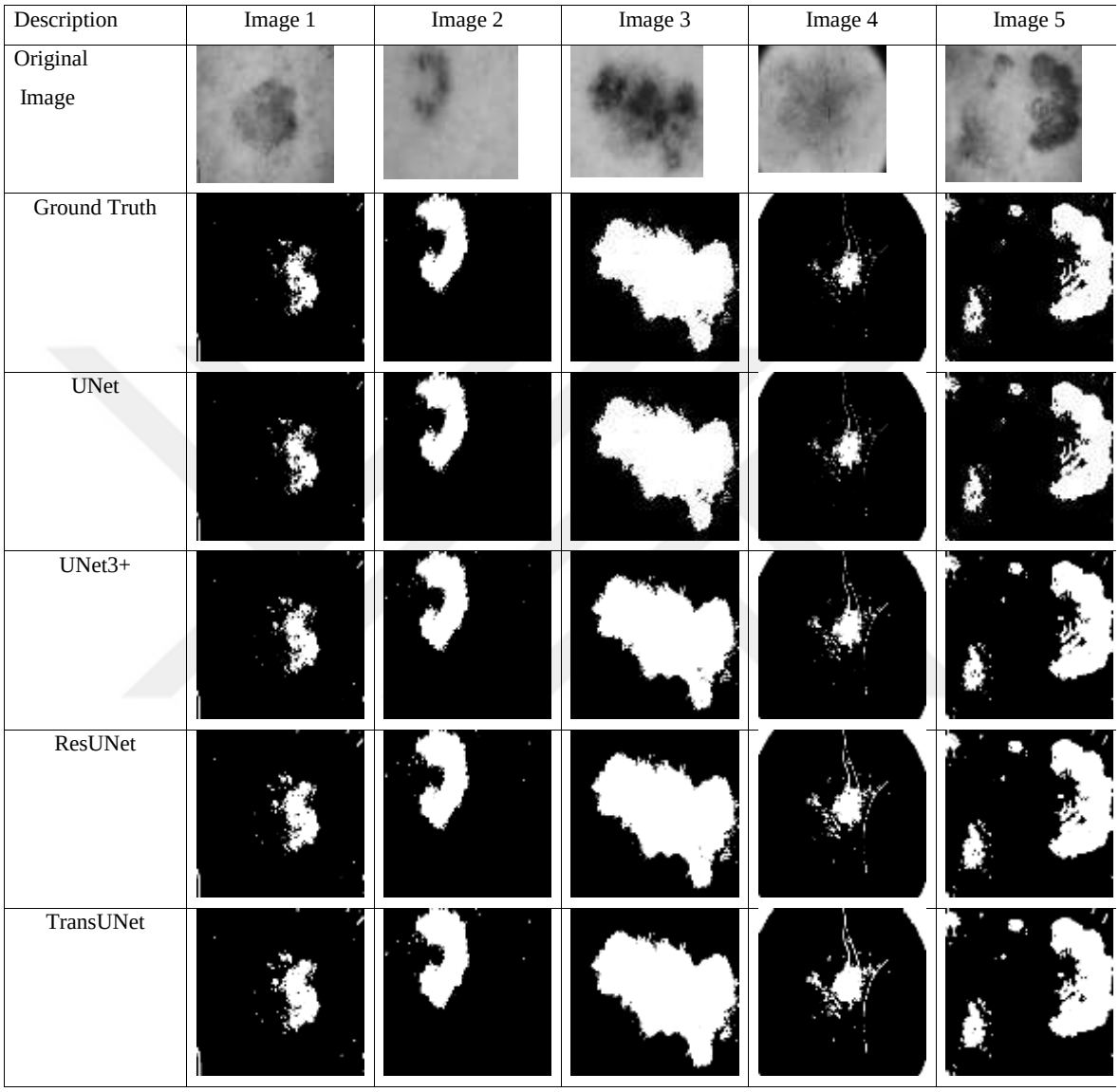


Figure 3.14: Segmentation Outcomes from the Executed ResDenseUnet++ Model.

3.7 PERFORMANCE EVALUATION OF THE DESIGNED IMAGE SEGMENTATION MODEL

The performance evaluation of the designed image segmentation model when compared with other conventional segmentation techniques is illustrated in Figure 14. The performance of the designed ResDenseUNet++-based segmentation model is evaluated in terms of

validation measures such as “accuracy, Jaccard, and dice coefficient”. The performance of the designed ResDenseUNet++-based segmentation model is evaluated for each metric while considering the statistical measures like the “standard deviation, mean, worst, best, and median”. Except for the standard deviation, the performance offered by the designed ResDenseUNet++-based segmentation model is higher. Out of all the statistical measures, enhanced performance is provided by the designed ResDenseUNet++-based segmentation model when best is taken as the statistical measure. When best is taken as the statistical measure, the accuracy of the designed ResDenseUNet++-based segmentation model is 2.13%, 4.35%, 6.67%, and 7.26% more than the TransUNet, ResUNet, UNet3+, and UNet models, respectively. Thus, it is proved that enriched performance on the segmentation of the skin cancer-based dermoscopy images is made possible by the designed ResDenseUNet++-based segmentation model when compared to the other conventional segmentation techniques.

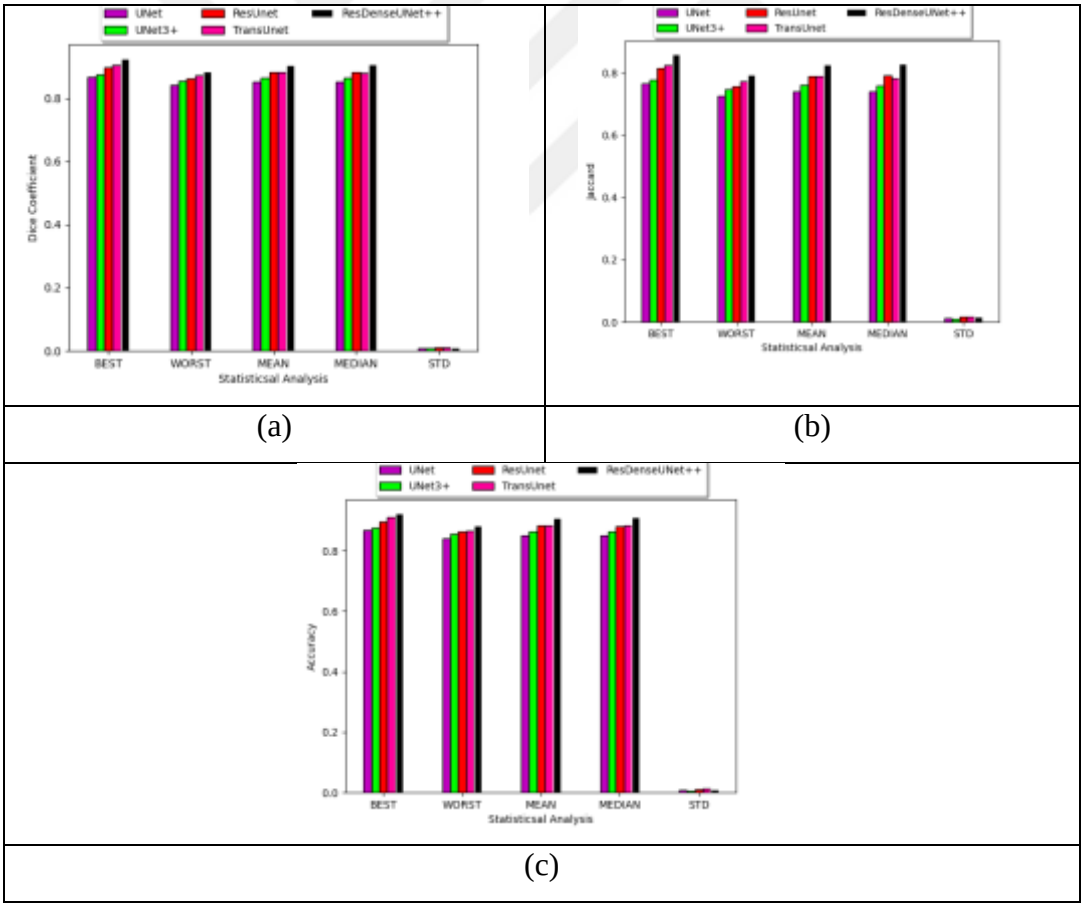


Figure 3.15: Performance Evaluation of the Designed Image Segmentation Model in Terms of “(a) Dice Coefficient, (b) Jaccard, and (c) Accuracy”.

3.8 ROC ANALYSIS OF THE SUGGESTED SKIN CANCER DETECTION SYSTEM

The ROC analysis of the suggested skin cancer detection system is shown in Figure 15. The ROC analysis is carried out by varying the "True Positive Rate (TPR)" against the FPR. The AUC of the suggested ViT-ShuffleNetV3-based skin cancer detection model is much more than that of the ShuffleNet, ResNet, RAN, and CNN models. When considering the FPR value as 0.2, the TPR of the suggested ViT-ShuffleNetV3-based skin cancer detection model is 1.08%, 4.44%, 6.82%, and 11.9% improved than the ShuffleNet, ResNet, RAN, and CNN models, respectively.

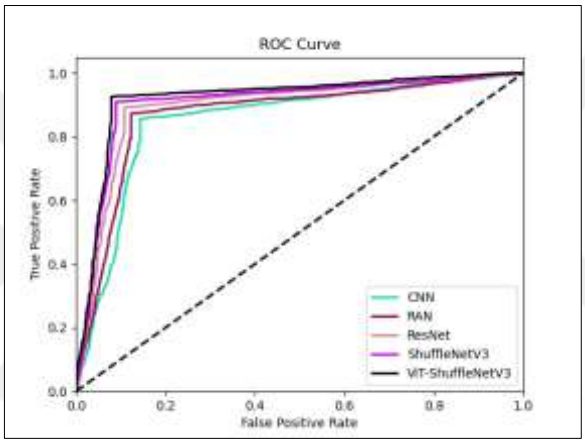


Figure 3.16: ROC Analysis of the Suggested Skin Cancer Detection System.

3.9 PERFORMANCE VALIDATION OF THE DEPLOYED SKIN CANCER DETECTION PROTOTYPE

The performance validation of the deployed skin cancer detection prototype when varying the learning rate and the batch size is given in Table 3. The deployed ViT-ShuffleNetV3-based skin cancer detection model is 53.63%, 47.27%, 35.16%, and 23.74% more precise than the CNN< RAN, ResNet, and ShuffleNetV3 models when considering the learning rate as 75%. Hence enhanced detection of skin cancer by the deployed ViT-ShuffleNetV3-based skin cancer detection model is verified.

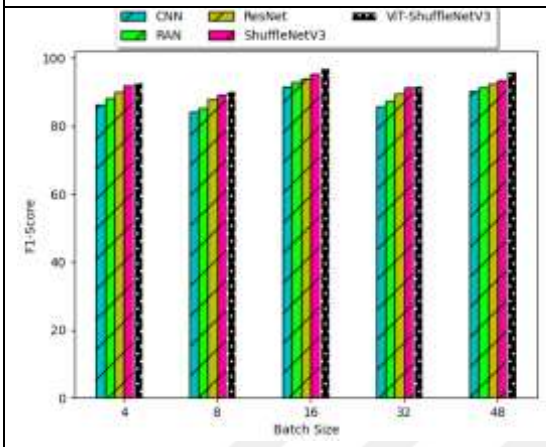
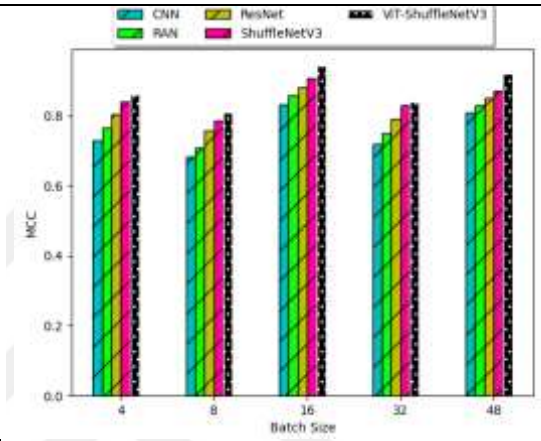
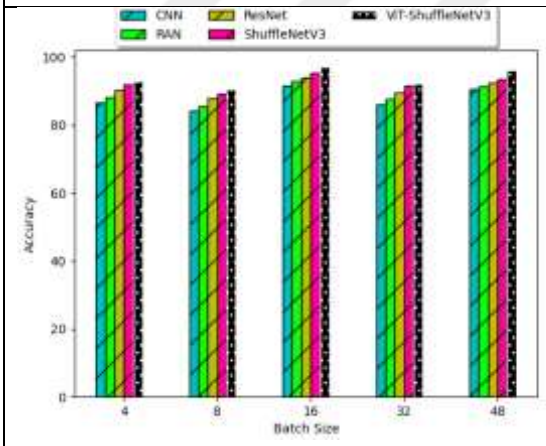
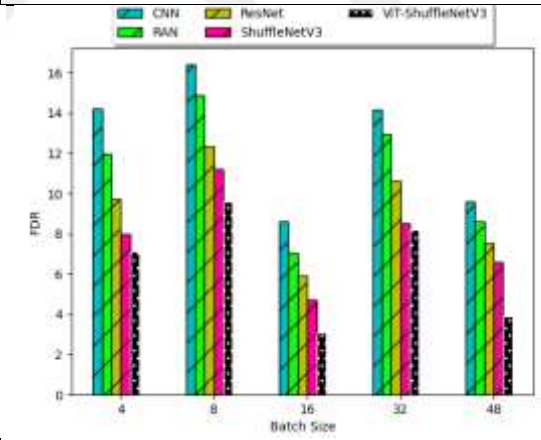
Table 3.2: Performance Validation of the Deployed Skin Cancer Detection Prototype.

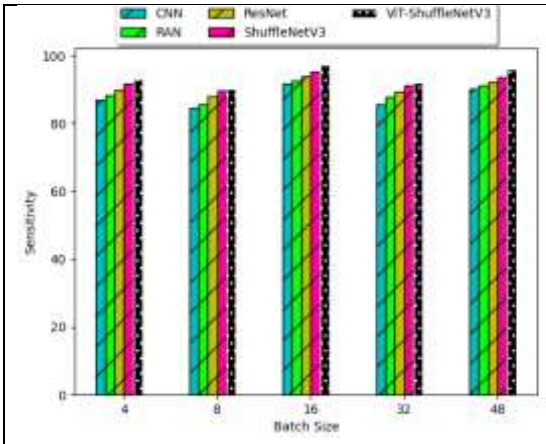
Terms/ Models	CNN [47]	RAN [48]	ResNet [49]	ShuffleNetV3	ViT-ShuffleNetV3
Batch Size-based Performance Validation					
"NPV"	85.884	88.259	90.501	92.216	93.008
"Sensitivity"	87.062	88.275	89.892	91.779	92.992
"FNR"	12.938	11.725	10.108	8.221	7.008
"F1-Score"	86.421	88.156	90.074	91.903	92.929
"FPR"	14.116	11.741	9.499	7.784	6.992
"Specificity"	85.884	88.259	90.501	92.216	93.008
"FDR"	14.210	11.962	9.743	7.973	7.133
"Accuracy"	86.467	88.267	90.200	92.000	93.000
"MCC"	0.729	0.765	0.804	0.840	0.860
"Precision"	85.790	88.038	90.257	92.027	92.867
Learning Rate-based Performance Validation					
"Accuracy"	89.057	89.701	90.950	92.160	94.782
"FPR"	10.942	10.295	9.046	7.837	5.215
"MCC"	0.529	0.546	0.580	0.616	0.709
"Specificity"	89.058	89.705	90.954	92.163	94.785
"NPV"	89.058	89.705	90.954	92.163	94.785
"Sensitivity"	89.046	89.656	90.904	92.122	94.738
"FNR"	10.954	10.344	9.096	7.878	5.262
"FDR"	63.240	61.651	58.216	54.358	43.524
"F1-Score"	52.038	53.720	57.252	61.041	70.766
"Precision"	36.760	38.349	41.784	45.642	56.476

3.10 FUNCTIONING EXAMINATION OF THE IMPLEMENTED SKIN CANCER DETECTION FRAMEWORK

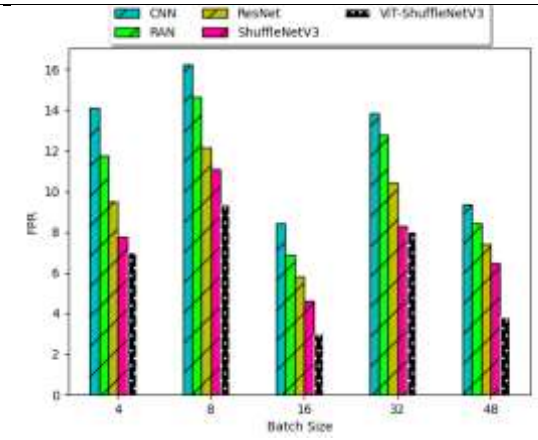
The functioning examination of the implemented skin cancer detection framework when varying the batch size and the learning rate is shown in Figure 16. The implemented ViT-ShuffleNetV3-based skin cancer detection model is 2.13%, 4.38%, 6.67%, and 9.09% more accurate than the ShuffleNetV3, ResNet, RAN, and CNN models, when the batch size is 16. The best performance is offered by the implemented ViT-ShuffleNetV3-based skin cancer detection model when the batch size reaches 16 regarding all the validation metrics. Similarly, the implemented ViT-ShuffleNetV3-based skin cancer detection model is 6.22%,

8.28%, 10.82%, and 13.22% more accurate than the ShuffleNetV3, ResNet, RAN, and CNN models, when the learning rate is 45. It is observed that as the learning rate increases, the performance of the implemented ViT-ShuffleNetV3-based skin cancer detection model also enhances with respect to every validation measure. Thus, it is proven that the implemented ViT-ShuffleNetV3-based skin cancer detection model gives better performance in the detection of skin cancer than other existing detection models.

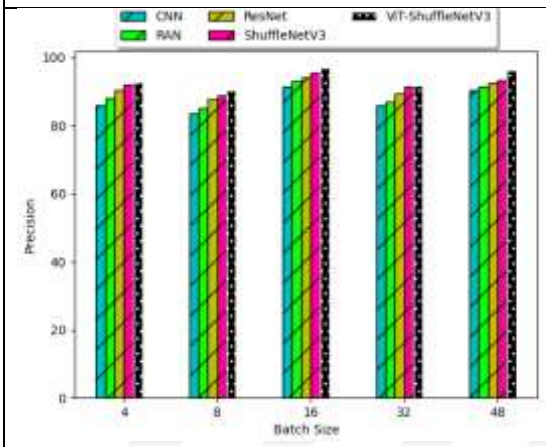
<table>
<tr><td colspan="2">Batch Size-based Examination</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>(a)</td><td>(b)</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>(c)</td><td>(d)</td></tr>
</table>



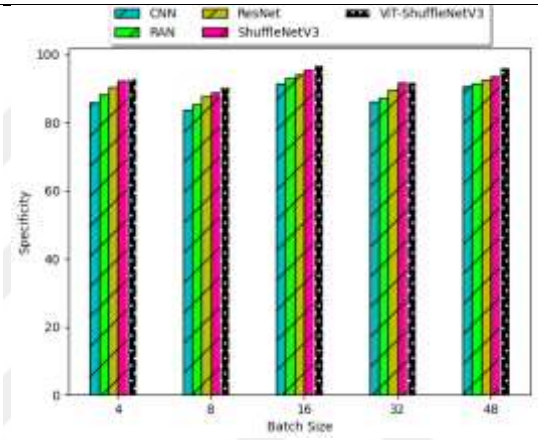
(e)



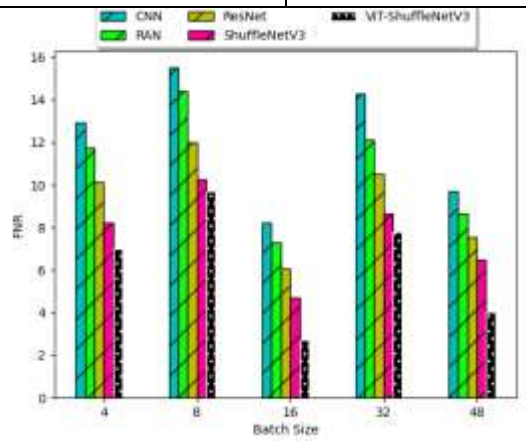
(f)



(g)

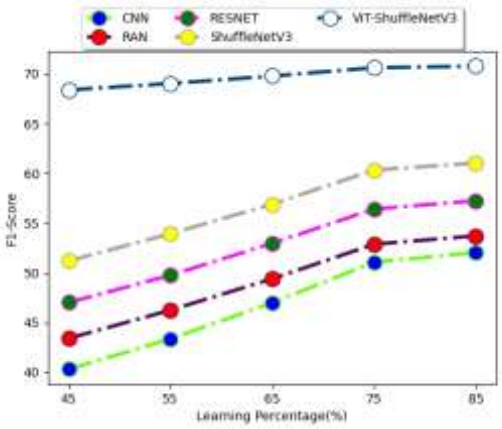


(h)

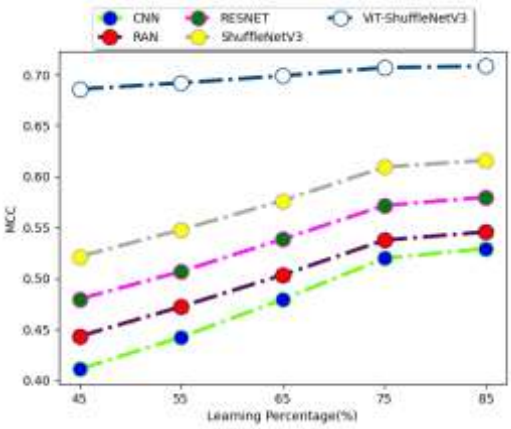


(i)

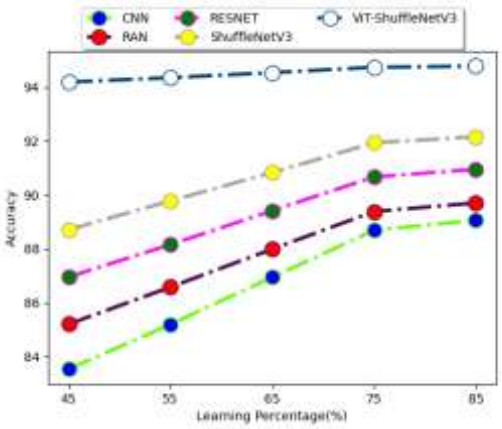
Learning Rate-based Examination


CNN
RESNET
VIT-ShuffleNetV3
RAN
ShuffleNetV3
F1-Score
Learning Percentage(%)

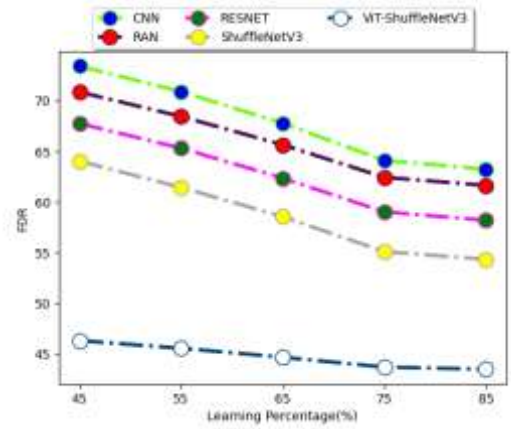
(a)


CNN
RESNET
VIT-ShuffleNetV3
RAN
ShuffleNetV3
MCC
Learning Percentage(%)

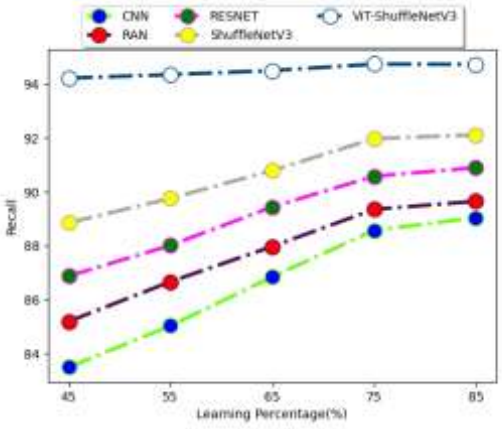
(b)


CNN
RESNET
VIT-ShuffleNetV3
RAN
ShuffleNetV3
Accuracy
Learning Percentage(%)

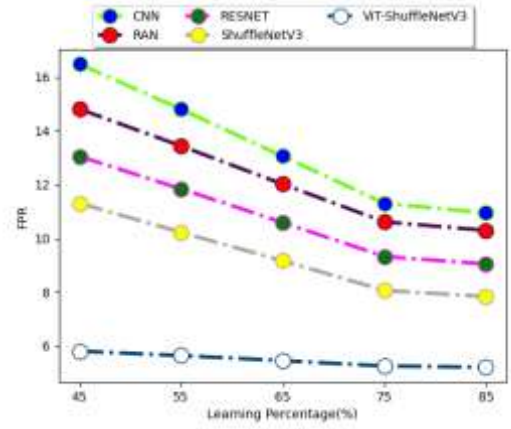
(c)


CNN
RESNET
VIT-ShuffleNetV3
RAN
ShuffleNetV3
FDR
Learning Percentage(%)

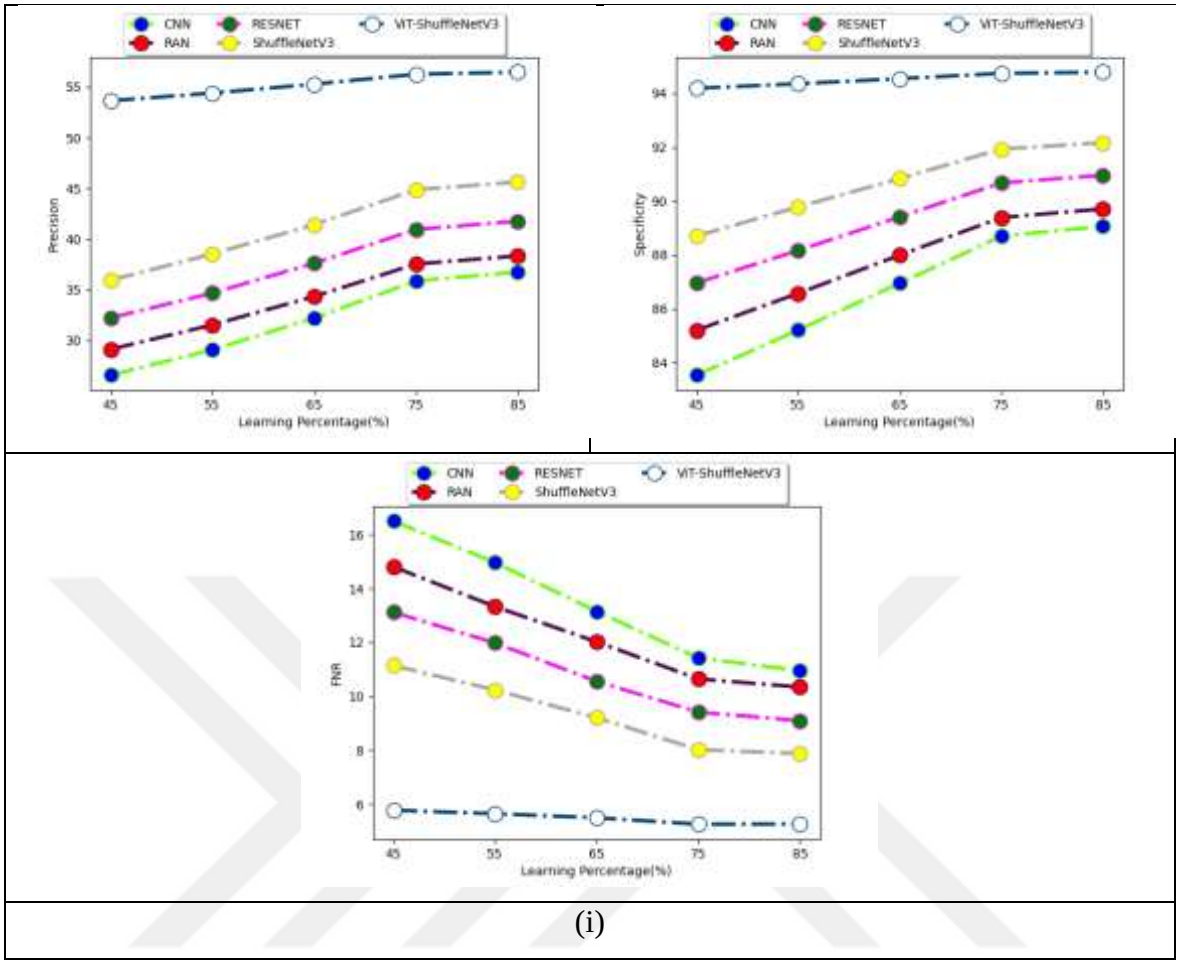
(d)


CNN
RESNET
VIT-ShuffleNetV3
RAN
ShuffleNetV3
Recall
Learning Percentage(%)

(e)


CNN
RESNET
VIT-ShuffleNetV3
RAN
ShuffleNetV3
FPR
Learning Percentage(%)

(f)



(i)

Figure 16: Functioning Examination of the Implemented Skin Cancer Detection Framework in terms of "(a) F1-score, (b) MCC, (c) Accuracy, (d) FDR, (e) Sensitivity, (f) FPR, (g) Precision, (h) Specificity, and (i) FNR".

4. CONCLUSION

An efficient skin cancer detection framework using deep learning techniques to resolve several complications in the classical techniques was developed in this work. The major phases of the implemented skin cancer detection framework were image collection, image segmentation, and skin cancer detection. Initially, the essential images utilized for the experiments were attained from the benchmark resources. Next, the collected features were provided as the input to the segmentation phase. Then, effective segmentation was performed using the ResDenseunet++ model. Here, the developed ResDenseunet++ provided the final segmented image as the outcome. Moreover, texture pattern images and gradient filter images were attained from segmented images. Further, the attained image outcomes were used as the input in the skin cancer detection phase. In this phase, effective skin cancer recognition was performed using developed ViT-ShufflenetV3. Experimentation was carried out to prove the efficacy of the implemented skin cancer detection model. Experimental outcomes showed that the accuracy offered by the suggested Vit-ShuffleV3 was 2.13%, 4.38%, 6.67%, and 9.09% enhanced than the ShuffleNetV3, ResNet, RAN, and CNN models when the batch size is 16. Thus, the implemented skin cancer detection framework was a secure highly accurate disease-classified outcome than the conventional techniques in different experimental validations.

4.1 FUTURE SCOPE

Some of the modifications and improvements that can be made for this model in the future are provided below.

a. As there is no extensive database with skin cancer-related data used in this work, the implemented deep learning model may show a limited performance which is affected by the small size of the dataset. Thus, a model that will be trained with a wide range of data will be executed in the future.
b. The majority of images regarding skin cancer used in this work as the dataset contain images of white skin. Testing the implemented deep learning system on such a skin tone alone results in the reduction of the overall accuracy of the model and makes it suffers from generalization issue. Thus, in the future, this model will be trained with dataset

that alleviate the issues with respect to the color bias in skin by gathering data with different skin tones.

c. As the input image has noise in it and the presence of complicated background is a bothering issue, suitable pre-processing procedure will be adopted in the near future.
d. The presence of hair in the skin can affect the detection performance of the suggested model. Therefore, a technique to remove the presence of hair from the provided input image will be considered as the future work.

REFERENCES

[1] Duong, Linh T, Nhi H Le, Toan B Tran, Vuong M Ngo, and Phuong T. Nguyen, "Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning," Expert Systems with Applications, vol. 184, no.115519, 2021.

[2] Dascalu A, and E. O. David, "Skin cancer detection by deep learning and sound analysis algorithms: a prospective clinical study of an elementary dermoscope," EBioMedicine, vol. 43, pp. 107–113, 2019.

[3] Wei, Lisheng, Kun Ding, and Huosheng Hu, "Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network," IEEE Access, vol. 8, pp. 99633-99647, 2020.

[4] Nida, Nudrat, Aun Irtaza, Ali Javed, Muhammad Haroon Yousaf, and Muhammad Tariq Mahmood, "Melanoma lesion detection and segmentation using deep region-based convolutional neural network and fuzzy C-means clustering," International journal of medical informatics, vol. 124, pp. 37-48, 2019.

[5] Kadampur, Mohammad Ali, and Sulaiman Al Riyaee, "Skin cancer detection: Applying a deep learning-based model driven architecture in the cloud for classifying dermal cell images," Informatics in Medicine Unlocked, vol. 18, no. 100282, 2020.

[6] Venugopal, Vipin, Navin Infant Raj, Malaya Kumar Nath, and Norton Stephen, "A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images," Decision Analytics Journal, no.100278, 2023.

[7] Lembhe, Ashutosh, Pranav Motarwar, Rudra Patil, and Susan Elias, "Enhancement in Skin Cancer Detection using Image Super Resolution and Convolutional Neural Network," Procedia Computer Science, vol. 218, pp.164-173, 2023.

[8] Adla, Devakishan, G. Venkata Rami Reddy, Padmalaya Nayak, and G. Karuna, "Deep learning-based computer-aided diagnosis model for skin cancer detection and classification," Distributed and Parallel Databases, vol. 40, no. 4, pp.717-736, 2022.

[9] Nawaz, Marriam, Zahid Mehmood, Tahira Nazir, Rizwan Ali Naqvi, Amjad Rehman, Munwar Iqbal, and Tanzila Saba, "Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering," Microscopy research and technique, vol. 85, no. 1, pp. 339-351, 2022.

[10] Davis, Lauren E, Sara C. Shalin, and Alan J. Tackett, "Current state of melanoma diagnosis and treatment," Cancer biology & therapy, vol. 20, no. 11, pp. 1366-1379, 2019.

[11] Adegun, Adekanmi A, and SerestinaViriri, "FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images," IEEE Access, vol. 8, pp. 150377-150396, 2020.

[12] Srividhya, V, K. Sujatha, R. S. Ponmagal, G. Durgadevi, and L. Madheshwaran, "Vision-based detection and categorization of skin lesions using deep learning neural networks," Procedia Computer Science, vol. 171, pp. 1726-1735, 2020.

[13] Wang, Yuheng, Daniel C. Louie, Jiayue Cai, Lioudmila Tchvialeva, Harvey Lui, Z. Jane Wang, and Tim K. Lee, "Deep learning enhances polarization speckle for in vivo skin cancer detection," Optics & Laser Technology, vol. 140, no.107006, 2021.

[14] Nahata, Hardik, and Satya P. Singh "Deep learning solutions for skin cancer detection and diagnosis," Machine Learning with Health Care Perspective: Machine Learning and Healthcare, pp. 159-182, 2020.

[15] Jayapriya, Kalyanakumar, and Israel Jeena Jacob, "Hybrid fully convolutional networks-based skin lesion segmentation and melanoma detection using deep feature," International Journal of Imaging Systems and Technology, vol. 30, no. 2, pp. 348-357, 2020.

[16] Pacheco, Andre GC, and Renato A. Krohling, "Recent advances in deep learning applied to skin cancer detection," arXiv preprint arXiv, pp. 1912.03280, 2019.

[17] Mazhar, Tehseen, Inayatul Haq, Allah Ditta, Syed Agha HassnainMohsan, Faisal Rehman, Imran Zafar, Jualang Azlan Gansau, and Lucky Poh Wah Goh, "The role of machine learning and deep learning approaches for the detection of skin cancer," Healthcare, vol. 11, no. 3, p. 415, 2023.

[18] Kumar, Manoj, Mohammed Alshehri, Rayed AlGhamdi, Purushottam Sharma, and Vikas Deep, "A de-ann inspired skin cancer detection approach using fuzzy c-means clustering," Mobile Networks and Applications, vol. 25, pp.1319-1329, 2020.

[19] Gerges, Firas, and Frank Y. Shih, "A convolutional deep neural network approaches for skin cancer detection using skin lesion images," International Journal of Electrical and Computer Engineering, vol. 15, no. 8, pp. 475-478, 2021.

[20] Tan, Teck Yan, Li Zhang, and Chee Peng Lim, "Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models," Applied Soft Computing, vol. 84, no.105725, 2019.

[21] Gomathi, E, M. Jayasheela, M. Thamarai, and M. Geetha, "Skin cancer detection using dual optimization based deep learning network," Biomedical Signal Processing and Control, vol. 84, no.104968, 2023.

[22] Agrahari, Pradhumn, Archit Agrawal, and N. Subhashini, "Skin cancer detection using deep learning," Futuristic Communication and Network Technologies, pp. 179-190, 2022.

[23] Narmatha, P., Shivani Gupta, TR Vijaya Lakshmi, and D. Manikavelan, "Skin cancer detection from dermoscopic images using Deep Siamese domain adaptation convolutional Neural Network optimized with Honey Badger Algorithm," Biomedical Signal Processing and Control, vol. 86, no. 105264, 2023.

[24] Akilandasowmya, G., G. Nirmaladevi, S. U. Suganthi, and A. Aishwariya, "Skin cancer diagnosis: Leveraging deep hidden features and ensemble classifiers for early detection and classification," Biomedical Signal Processing and Control, no. 105306, 2023.

[25] Gouda, Walaa, Najm Us Sama, Ghada Al-Waakid, Mamoona Humayun, and Noor Zaman Jhanjhi, "Detection of skin cancer based on skin lesion images using deep learning," Healthcare, vol. 10, no. 7, pp. 1183, 2022.

[26] Hosny, Khalid M., Mohamed A. Kassem, and Mohamed M. Fouad, "Classification of skin lesions into seven classes using transfer learning with AlexNet," Journal of digital imaging, vol. 33, pp. 1325-1334, 2020.

[27] Tabrizchi, Hamed, Sepideh Parvizpour, and Jafar Razmara, "An improved VGG model for skin cancer detection," Neural Processing Letters, vol. 55, no. 4, pp. 3715-3732, 2023.

[28] Mohapatra, Subasish, N. V. S. Abhishek, Dibyajit Bardhan, Anisha Ankita Ghosh, and Shubhadarshini Mohanty, "Comparison of MobileNet and ResNet CNN Architectures in the CNN-Based Skin Cancer Classifier Model," Machine Learning for Healthcare Applications, pp. 169-186, 2021.

[29] Adegun, Adekanmi A., and Serestina Viriri, "FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images," IEEE Access, vol. 8, pp. 150377-150396, 2020.

[30] Widiansyah, M., S. Rasyid, P. Wisnu, and Adi Wibowo, "Image segmentation of skin cancer using MobileNet as an encoder and linknet as a decoder," Journal of Physics, vol. 1943, no. 1, p. 012113, 2021.

[31] Rehman, Hafeez Ur, Nudrat Nida, Syed Adnan Shah, Wakeel Ahmad, Muhammad Imran Faizi, and Syed Muhammad Anwar, "Automatic melanoma detection and segmentation in dermoscopy images using deep RetinaNet and conditional random fields," Multimedia Tools and Applications, vol. 81, no. 18, pp. 25765-25785, 2022.

[32] Lee, Bumshik, Nagaraj Yamanakkanavar, and Jae Young Choi, "Automatic segmentation of brain MRI using a novel patch-wise U-net deep architecture," Plos one 15, no. 8, e0236493, 2020.

[33] Zhou, Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "Unet++: A nested u-net architecture for medical image segmentation," Springer International Publishing, pp. 3-11, 20 September 2018.

[34] Shehab, Lamia H., Omar M. Fahmy, Safa M. Gasser, and Mohamed S. El-Mahallawy, "An efficient brain tumor image segmentation based on deep residual networks (ResNets)," Journal of King Saud University-Engineering Sciences, vol. 33, no. 6, pp. 404-412, 2021.

[35] Zhang, Yulun, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, "Residual dense network for image super-resolution," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2472-2481, 2018.

[36] Gong, Cheng, Jing Liu, Ming Gong, Jingbing Li, Uzair Aslam Bhatti, and Jixin Ma. "Robust medical zero-watermarking algorithm based on Residual-DenseNet." IET Biometrics, vol. 11, no. 6, pp. 547-556, 2022.

[37] W. Shi, J. Caballero, F. Husz´ar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," CVPR, 2016.

[38] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," CVPR, 2017.

[39] Huruy Tesfai, Hani Saleh, Mahmoud Al-Qutayri, Moath B. Mohammad, Temesghen Tekeste, Ahsan Khandoker, and Baker Mohammad, "Lightweight ShufflenetBased CNN for Arrhythmia Classification," IEEE Access, vol. 10, pp. 111842-111854, 2022.

[40] Yixian Fu, Yuanyao Lu and Ran Ni, "Chinese Lip-Reading Research Based on ShuffleNet and CBAM," Applied Science, vol. 13, no. 2, pp. 1106, 2023.

[41] Zhang, Xiangyu, Xinyu Zhou, Mengxiao Lin, and Jian Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6848-6856, 2018.

[42] Usman, Mohammad, Tehseen Zia, and Ali Tariq, "Analyzing transfer learning of vision transformers for interpreting chest radiography," Journal of digital imaging, vol. 35, no. 6, pp. 1445-1462, 2022.

[43] Guan, Steven, Amir A. Khan, Siddhartha Sikdar, and Parag V. Chitnis, "Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal," IEEE journal of biomedical and health informatics, vol. 24, no. 2, pp. 568-576, 2019.

[44] Yin, Meijie, Peng Wang, Cui Ni, and Weilong Hao, "Cloud and snow detection of remote sensing images based on improved Unet3+," Scientific Reports 12, no. 1, pp. 14415, 2022.

[45] Lan, Yancheng, and Xuming Zhang, "Real-time ultrasound image despeckling using mixed-attention mechanism based residual UNet," IEEE Access, vol. 8, pp. 195327-195340, 2020.

[46] Chen, Jieneng, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint, 2021.

[47] Zhang, Ni, Yi-Xin Cai, Yong-Yong Wang, Yi-Tao Tian, Xiao-Li Wang, and Benjamin Badami, "Skin cancer diagnosis based on optimized convolutional neural network," Artificial intelligence in medicine, vol. 102, pp. 101756, 2020.

[48] Ling, Hefei, Jiyang Wu, Lei Wu, Junrui Huang, Jiazhong Chen, and Ping Li, "Self-residual attention network for deep face recognition," IEEE Access, vol. 7, pp. 55159-55168, 2019.

[49] Zhang, Ke, Miao Sun, Tony X. Han, Xingfang Yuan, Liru Guo, and Tao Liu, "Residual networks of residual networks: Multilevel residual networks," IEEE

Transactions on Circuits and Systems for Video Technology, vol. 28, no. 6, pp. 1303-1314, 2017.