Enhancing Skin Cancer Detection: A Comparative Analysis of Models with VGG-16, VGG-19, and Inception V3

Kiran Likhar [1], Dr. Sonali Ridhorkar [2]

Abstract: In the realm of skin cancer-related fatalities, early detection of malignant lesions is the key to effective treatment and saving lives. While deep learning approach have shown promise in cancer detection, the effectiveness of individual models can be limiting. In this article, we explore the potential of ensemble models to enhance the performance of skin cancer detection. We present an ensemble model designed to identify skin cancer, leveraging the power of three well-established deep learning design: VGG-16, VGG-19, and Inception V3. By comparing the performance of these models, we shed light on their strengths and weaknesses in this critical domain. Our findings reveal that the suggested ensemble model, with a particular emphasis on VGG-16, exhibits an impressive average accuracy of 92%. Notably, when compared to VGG-19 and Inception V3, the suggested VGG-16 model outperforms in various crucial aspects. It excels in terms of sensitivity, accuracy, F-Score, specificity, false-positive rate, and precision, create it a promising choice for accurate and genuine skin cancer detection. In the pursuit of improving early cancer diagnosis, this research underscores the potential of ensemble models and highlights the pivotal role played by the VGG-16 architecture. These results provide valuable insights for the medical community and deep learning practitioners, with the ultimate goal of enhancing skin cancer detection methods and saving lives.

1. Introduction

Cells in the human body normally replicate in a predictable manner. To ensure proper bodily function, its birth, effective phase, and death should take place in the correct order. The disruption of the natural order causes the emergence of numerous disorders, including cancer. Anywhere among the billions of cells that make up the human structure, Cancer can occur. When a human being gets cancer, certain biological cells initiate to divide uncontrollably and disperse into the tissues around them. Human cells typically divide and increase in number to create new cells as needed by the body. The method causes cells to develop, age, or get tainted; as a result, new, healthy cells grow in its place after it dies. Once cancer takes hold, the methodical, precise cell-breaking mechanism is destroyed. The threshold for cell irregularity and damage consequently rises dramatically. New tissues are only created when they are necessary, and cell viability only occurs when the previous ones die. If the embryonic cells are not needed, they will continue to divide unabated and might lead to the development of a tumour [1].

Both benign and malignant tumours are brought on by cancer. Malignant tumours are those that primarily contain cancerous cells. Malignancy denotes a cell's ability to invade or spread to nearby tissues. When malignant tumours divide, a few tumour-causing cells travel via the blood or lymph system to distant areas of the body, where they form secondary tumours that are separated from the original tumour. Benign tumours do not divide inside the body or seize tissues, in comparison to malignant tumours. It is noted that benign tumours typically have a larger size. However, once benign tumours are take away, they cannot come back, although malignant tumours may do so following surgery. The benign tumour in the brain poses a serious constraint to life, in contrast to many benign tumours throughout the body that are often not harmful [2]. Figures 1 and 2 display examples of benign and malignant dermoscopy images, respectively.
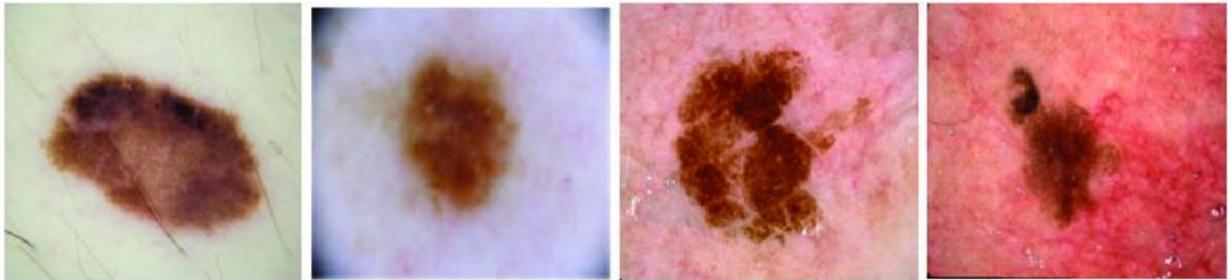
[1] *PhD Research Scholar G H Raisoni University Amravati , India*
[2] *Associate Professor G H Raisoni Institute of Engineering and Technology, India*
kiran.likhar@raisoni.net

Fig 1: Malignant skin lesion diagnosis.

Fig 2: Benign dermoscopic image sample.

The human body can be shielded by the skin from UV rays, heat, damage, and bacterial or viral diseases. The skin stores fat and water while also assisting the body in regulating body temperature. Skin cancer, an extremely frequent kind of cancer, is regarded as the biggest problem in terms of public health. Skin cancer can begin anywhere on a human's skin, although it typically begins on skin that has been exposed to sunlight. There are several skin layers. However, skin cancer often starts in the epidermis layer, which is one of the outermost layers.

Non-melanoma skin tumours, which include basal cells and squamous cells, are among the many diverse types of skin cancer. Non-melanoma skin cancer responds closely to therapy and seldom widen to other parts of the body. Among the various skin cancer, melanoma is the most threatening. Malignant skin lesions come in two different varieties: melanocytic lesions, like melanoma, and nonmelanocytic lesions, such as basal cell carcinoma. Melanoma is the mass severe, aggressive, but least common kind of skin cancer [2]. If skin cancer caused by melanoma is not detected in a timely manner, it will likely spread to other regions of the body and engulf nearby tissues. Each year, there is a rise in the number of melanoma cases. 9730 fatalities from melanoma were anticipated in the United States, according to the Melanoma Foundation [3], a renowned cancer institute. Furthermore, it predicted a rise of 87,110 documented cases, or 200%, since 1973.

In the literature, machine learning and deep learning operated to build process for detecting cancer using protein sequences and image data. Machine learning algorithms require behaviours that have been human-engineered. The lengthy, labor-intensive, and subjective manual feature extraction process depends on the expert's subjective judgment. Deep learning techniques have somewhat solved this problem. Deep learning techniques enable automated feature engineering. Deep convolutional neural networks (DCNN) [4], which have been more popular recently because of this characteristic, are being used by academics to tackle issues across various fields, including the categorization of medical images. To increase classification performance, however, ensemble learning techniques have recently been proposed [5], [6], and [7].

The goal of transfer learning is to apply information acquired while resolving one problem to another that is related to it [8]. Additionally, Individual learning is limited to choosing between delicate options like cancer detection. The solution to this issue can be found by integrating the choices made by several students. In comparison to the individual learners, the homogenize choice is anticipated to be correct. Merging the decisions of separate learners can improve the accuracy of skin cancer diagnosis. The work that is being presented has used an ensemble models like VGG16, VGG19, and InceptionV3. The outcomes show that the VGG-16 ensemble model outperformed than the VGG-19 and Inception V3.

Further paper is divided into the following sections: Section II depict related work; Section III describes datasets; Section IV describes the proposed system;

Section V presents results and discussion; and Section VI reviews the conclusion and future work.

2. Literature Survey

Detecting skin cancer is a critical aspect of early detection and therapy, and extensive research has been conducted to improve detection procedure. The detection of skin cancer has made extensive use of machine learning and deep learning approach. By removing the manual elements from dermoscopy pictures, machine learning approach detect skin lesions. The author, Waheed et al., developed a machine learning technique for melanoma detection in dermoscopic images [9]. By using distinguishing data from altered genes in protein amino acid patterns, Mohsin et al. conducted cancer diagnosis [10]. A cancer prediction method was created by Abdul M. et al., utilising closest neighbour and SVM [11]. A support vector machine [12] has been used to identify melanoma [13]. They classified cancer using segmented pictures. These ML methods are constrained by the knowledge of dermatologists and need handmade characteristics.

Deep learning, AI-driven systems, and non-invasive imaging techniques have all contributed to improving the accuracy and accessibility of skin cancer diagnosis, potentially transfigure the field of dermatology. Utilising deep learning technique, mostly convolutional neural networks (CNNs), is one popular strategy, which have shown remarkable promise in automating skin cancer detection. Esteva et al. presented A significant improvement in the field of skin cancer detection by leveraging deep neural networks, more precisely convolutional neural networks (CNNs) [14]. They trained a CNN model using a vast dataset of dermoscopic images, which are high-resolution images of skin lesions captured through a specialized tool. The results were astonishing, as the CNN achieved classification accuracy on par with that of dermatologists. The study demonstrated the potential of artificial intelligence to provide reliable and rapid skin cancer diagnosis, with implications for improving patient outcomes through early detection and intervention. Haenssle et al. organize a pivotal study that not only validated capabilities of CNNs in skin cancer diagnosis but also compared their performance directly to that of dermatologists [15]. Their research revealed that deep learning convolutional neural network could match diagnostic accuracy of a large group of dermatologists. This work underscored the potential of AI systems to assist healthcare professionals, reduce diagnostic errors, and provide accessible and consistent skin cancer screening, especially in areas with limited access to dermatological expertise. Tschandl et al. introduced a novel AI system called "HIDEX," which goes beyond traditional dermoscopic image analysis. HIDEX incorporates both dermoscopy and clinical information,

enhancing diagnostic accuracy [16]. This approach illustrates the power of combining different data sources to improve skin cancer detection, taking into account not only image features but also clinical context. The research represents a step forward in developing AI tools that can integrate various data streams for more comprehensive and precise medical diagnoses. Halicek et al. delved into hyperspectral imaging of skin, a technique that captures the spectral signatures of skin lesions [17]. This non-invasive method has shown promise in distinguishing between benign and malignant skin conditions. By analyzing the unique spectral characteristics of tissues, hyperspectral imaging provides valuable diagnostic information, expanding the toolkit for skin cancer detection beyond traditional visual assessments and imaging modalities. Rajadhyaksha et al. pioneered using confocal laser scanning microscopy to image human skin in real time [18]. This innovative approach provided high-resolution in real-time images of skin lesions, allowing dermatologists to examine cellular and structural details non-invasively. Their work contributed to precise diagnosis and laid the foundation for advanced imaging techniques that offer insights into skin conditions at the microscopic level.

Another notable development is the integration of smartphone apps equipped with AI algorithms for skin cancer detection, such as "SkinVision" and "DermEngine," which empower users to perform preliminary self-assessments [19]. These smartphone apps, SkinVision and DermEngine, represent the growing trend of making skin cancer detection accessible to the general public. They utilize AI algorithms to analyze images of skin lesions captured with smartphone cameras. Users can perform preliminary self-assessments, with the apps providing risk assessments and recommendations for further evaluation by healthcare professionals. These apps are a testament to the potential of AI-driven technology to empower individuals to take control of their health and encourage early detection and intervention for skin cancer.

Ensemble networks are a common technique used by researchers today to improve classification performance. Typically, each model is trained separately, and then predictions from several models are combined using maximum voting and stacking procedures to generate results. An ensemble network was suggested by Aboulmira et al. [20] for classifying skin lesions. The separate models are used to extract the features, and the several models are then integrated to increase the classication rate. The publicly accessible ISIC-2018 dataset has been used to test the suggested ensemble of seven predictors, which performs better than previous approaches. FixCaps, a more effective capsule network, has been utilised in [21] for the early identification of skin

cancer. In comparison in the direction of baseline CapsNet with a huge kernel size of 31*31, the suggested method gained a larger receptive field, which improved its detection efficiency while lowering computational overhead. By retaining the short and long-term correlations, Cao et al. [22] presented a special network inter pixel correlation learning (ICL) for the initial detection of skin lesions. The suggested Pyramid transformer inter-pixel correlations (PTIC) are used to record global information in the model's encoder-decoder architecture, and local neighbourhood metric learning (LNML) is used to improve local semantic correlations. With the use of a two-stage methodology, segmentation performance may be improved, as measured by public challenge datasets, by increasing both intra-class consistency and inter-class variance. Using dermoscopic pictures, Javaid et al. [23] presented segmentation and classification of skin cancer by machine learning. Image

is segmented using the OTSU thresholding method, and then a grey level co-occurrence matrix (GLCM), a histogram of oriented gradient (HoG), and colour characteristics are retrieved for use in the ML model classification. In a different ML-based study, skin imaging data was used to detect skin cancer [24]. First, the skin is identified by using a median filter, and then the skin has been segmented using a mean shift algorithm. Individual learners' performance only involves decision-making, but this limitation are solved by merging the individual learners' decisions.

3. Dataset

According to the survey, there are numerous datasets with images of both cancer and non-cancerous tissue that are available for additional research. The detailed information of various dataset is discussed in table 1 as below,

Table 1: Various DataSets for Skin Cancers

Dataset Name	Year	Number of Images	Image Types	Skin Lesion Types	Metadata Information
ISIC 2017	2017	13,000+	Dermoscopy	Melanoma, Nevus, Keratosis	Patient info, Image acquisition details, Diagnoses
ISIC 2019	2019	25,000+	Dermoscopy	Melanoma, Nevus, Keratosis	Clinical data, Diagnostic annotations, Image metadata
HAM10000	2018	10,000+	Clinical and Dermoscopic	Melanoma, Nevus, Keratosis	Patient characteristics, Diagnostic information, Lesion metadata
PH2 Dataset	2014	200+	Dermoscopy	Melanoma, Nevus, Keratosis	Clinical data, Image acquisition details, Ground truth
BCN20000	2017	20,000+	Dermoscopic and Clinical	Melanoma, Nevus, Keratosis	Patient data, Lesion annotations, Image characteristics
DermoFit	2010	1,300+	Dermoscopy	Melanoma, Nevus, Keratosis	Patient information, Diagnosis, Image metadata
SKINLIFE	2015	200+	Clinical and Dermoscopic	Melanoma, Nevus, Keratosis	Clinical data, Lesion characteristics, Diagnostic annotations

As per survey it has been found that The "ISIS-2017" and "ISIS-2019" datasets are referring to are datasets related to skin cancer research. These datasets were released as part of the International Skin Imaging Collaboration (ISIC) initiative, which aims to enhance the early diagnosis of melanoma and other skin diseases through the use of dermatology images. The dataset includes Melanoma, Nevus, Basal Cell Carcinoma, Actinic Keratosis, and Benign Keratosis categories of dataset. The brief description of these datasets:

a. ISIC 2017 Dataset: - The ISIC 2017 dataset containing high-quality dermoscopy images of various skin lesions, including melanoma, nevus, and seborrheic keratosis. It contains over 13,000 labeled images, with a focus on melanoma, a potentially deadly form of skin cancer [25]. Each image in the dataset is accompanied by metadata such as patient information, image acquisition details, and diagnostic annotations. Researchers and data scientists use this dataset for tasks like skin

lesion classification and melanoma detection using machine learning and computer vision techniques.

b. ISIC 2019 Dataset: This dataset is a continuation of the ISIC initiative, featuring a larger and more diverse set of skin images. It includes over 25,000 labeled images, making it one of the largest publicly available skin cancer image datasets. Like ISIC 2017 dataset, it contains various types of skin lesions, with a strong emphasis on melanoma images [26]. This dataset also provides detailed metadata for each image, including clinical information and diagnostic annotations. Researchers and healthcare professionals use these datasets to develop and evaluate algorithms for the automated detection and classification of skin lesions, aiding in initial diagnosis and treatment of skin cancer. They are valuable resources for training and testing ML models for skin cancer detection and related tasks.

Based on above datasets ISIC 2017 and ISIC 2019 datasets are utilized for proposed work.

4. Proposed Methodology

Figure 3 shows the proposed framework. The dataset used is different images from ISIC 2017 and ISIC 2019. Data preprocessing such as reshaping normalization is performed on dataset. Than dataset is splited into 80:20 ratio such as 80% data for training the model and 20% data for testing the model are used. Three different ensemble models are used for training like VGG-16, VGG-19 and inception V3. Various stages involved in model development is explained in detail as follow,
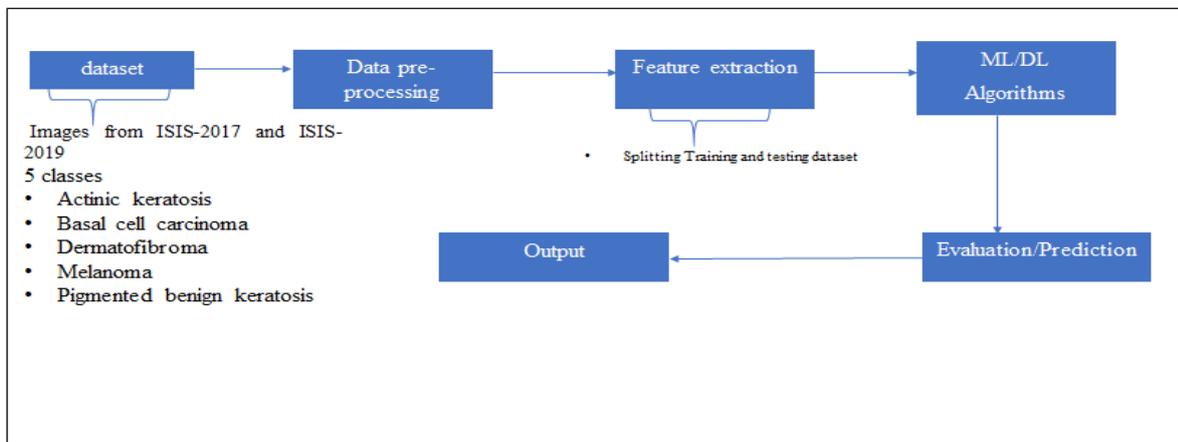

Fig 3. Proposed Framework

4.1 Preprocessing - In the data preprocessing phase for the skin cancer detection system, All images had to be altered to a standard resolution of 224 x 224 pixels, which was an important step. This uniform size ensures consistency across the dataset, simplifying the subsequent steps of model development and analysis. By resizing the images, we reduce computational complexity and create a foundation for efficient model training and evaluation. Furthermore, data normalization was performed on the cancer dataset. In order to do this, the image's pixel values had to be scaled to a standard variate, between 0 and 1. Normalization is essential for ensuring that the model converges faster during training and is not influenced by variations in pixel intensity among the images. These preprocessing steps collectively establish a solid groundwork for the development of a robust and accurate skin cancer detection model, enhancing its ability to make reliable predictions based on the resized and normalized data.3000 malignant and 2800 benign pictures were used to accomplish binary class classification utilising the ensemble technique that has been presented. Since there are only 2800 benign photos in this dataset, malignant images are divided by 2800 to keep the algorithm from being biassed. 80% of the total number of photos is used as training data, which makes up the dataset. The test data has been created using the remaining photos. Images of various sizes design the dataset. In the suggested technique, images have been downsized to 224x224x3.

4.2 Deep Neural Network - Three separate VGG 16, VGG 19, and Inception V3 convolution-based deep neural network models had been created in order to create suggested ensemble. The models' architectural progress is explained in detail below:

4.2.1 VGG 16: VGGNet is one of the most often applied CNN models. The VGG model's ease of use, its popularity as a deep learning model can be attributed to its clarity and utilisation of tiny convolutional kernels. The VGGNet architecture uses a 3x3 convolution kernel with max-pooling and ReLU layers, additionally three fully connected layers, for the extraction and classication of features. Smaller kernels are used in the creation, which results in some parameters and more effective training and testing. A series of 3x3-sized kernels may also be stacked to provide larger effective receptive fields (for example, 5x5 with two layers, 7x7 working with three layers. Much

importantly, smaller filters make it possible to stack more layers, which creates a deeper network, improves profitability on vision-related tasks. This clearly expresses the basic idea of the design, which promotes the usage of deeper networks for improved feature learning. After the input layer comes the VGG model layer, which is composed of five blocks. Proposed model reads pre-processed pictures with a size of 224x224 at the input layer using a VGG model. Following the input layer, the layer's first block begins with two convolutional layers, accompanied by the pooling layer. 64 filters make up the first convolutional layer. After polling, the final picture shrinks to 112x112 pixels in size. Each block of layers has

a nearly identical layering pattern. In the second block, features are shrunk to a size of 56x56, and the first and second convolutional layers each have 128 filters. The feature map is shrunk to 28x28 in the third block, which consists of three convolutional layers with 256 filters and max-pooling. The feature map is reduced to a size of 14x14 by the pooling layer in the fourth block, which also comprises three convolutional layers with 512 filters. The feature map is further condensed to 7x7 in the final, fifth block, which consists of 3 convolutional layers with 512 filters. Fig. 4 depicts the precise architecture of the VGG16 [27].
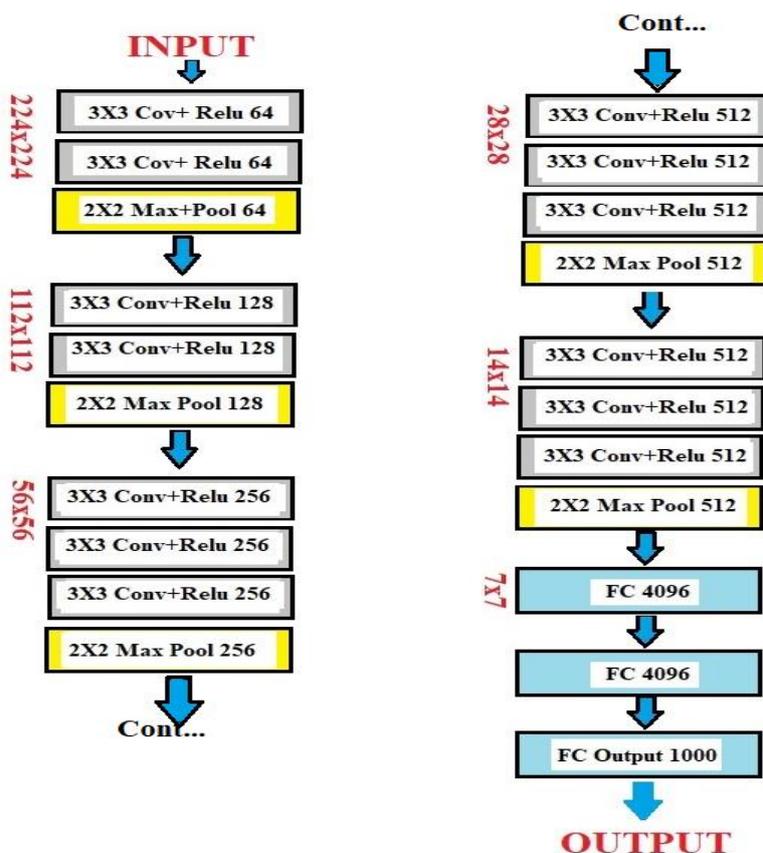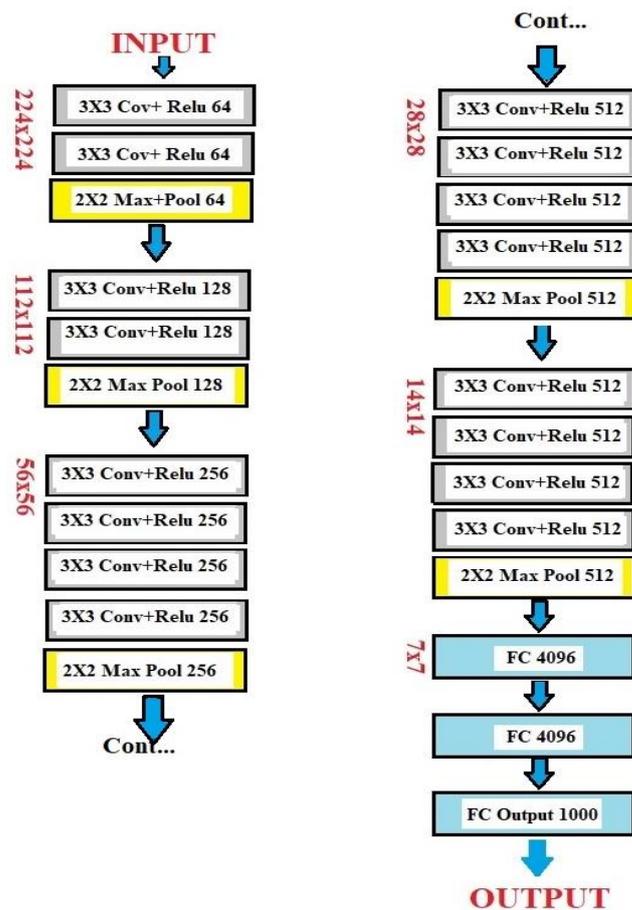

Fig. 4 VGG 16 based Cancer Detection System

4.2.2 VGG-19:- VGG19, an extension of the VGG16 model, is a significant advancement in deep learning for image analysis and classification tasks. Much like its predecessor, VGG19 is celebrated for its simplicity and practicality. It adopts the use of small 3x3 convolutional kernels, complemented by max-pooling layers and Rectified Linear Unit (ReLU) activations, which are the hallmark features of the VGGNet family. The rationale behind this design philosophy is to employ smaller filters that facilitate a deeper network, resulting in more effective feature learning. In the VGG19 architecture, the input layer is succeeded by five distinct blocks, each

characterized by a consistent layering pattern. As we progress through the blocks, the count of filters in each convolutional layer incremented, enhancing the network's capability to extract increasingly complex and abstract features. The final block further reduces the spatial dimensions of the feature maps, accompanied in a highly compact representation of the input image. VGG19's deep and consistent architecture underlines the notion that deeper networks excel in vision-related tasks, further solidifying the legacy of the VGGNet family within the field of deep learning for computer vision. Fig. 5 depicts precise architecture of the VGG19 [28].

Fig. 5 VGG 19 based Cancer Detection System

4.2.3 InceptionV3:- The Inception v3 architecture has proven to be highly effective in the realm of computer vision, and it stands as a robust foundation for developing a skin cancer detection system. In this specific implementation, the model is tailored to work with images of size 224x224x3, which is a common resolution for many deep learning applications. The Inception v3 architecture, characterized by its multiple branches and the efficient use of various filter sizes within its inception modules, is employed to capture intricate features in skin lesion images. The multi-scale approach allows the model to detect patterns of varying sizes and complexities, which is crucial for identifying potential malignancies. The System is trained on a dataset of skin images, which had been pre-processed to ensure consistency and accuracy. In this skin cancer detection system, Inception v3 excels at

extracting relevant features from the images, facilitating the differentiation between benign and malignant lesions. The deep layers of the network, along with the residual connections, enable it to capture subtle patterns that may indicate the presence of cancerous cells. The model is trained to categorize skin lesions as either benign or malignant, providing a valuable tool for early diagnosis and intervention. With its deep architecture and multi-scale feature extraction, Inception v3 proves to be a powerful asset in the ongoing battle against skin cancer. It's worth noting that in practice, clinical validation and continuous model improvement are essential to ensure the system's effectiveness in real-world healthcare scenarios. Fig. 6 depicts the precise architecture of the InceptionV3 [29].
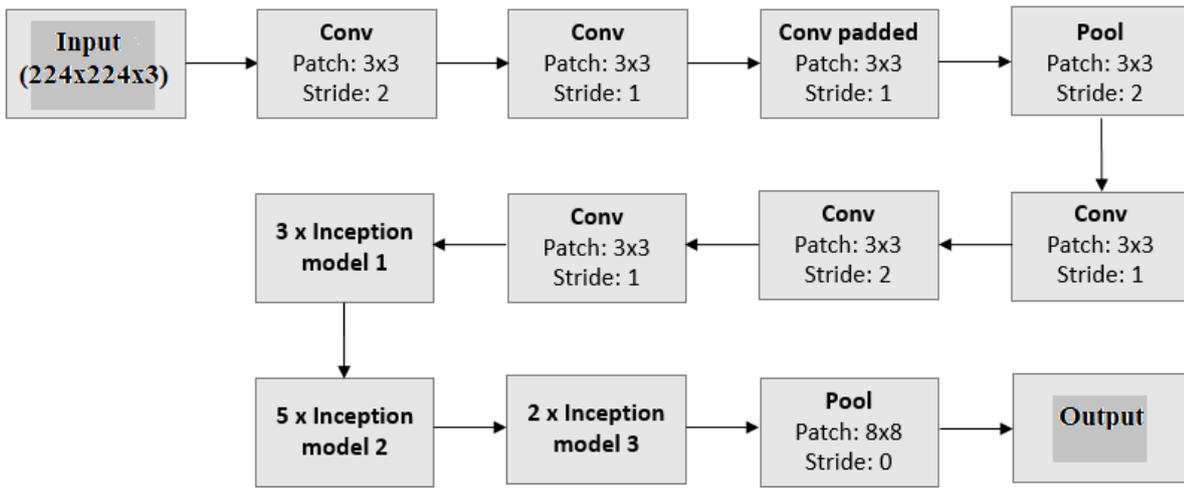
Fig 6. InceptionV3 based Cancer Detection System

5. Results and Discussion

5.1 Factors Used to Measure Performance

The effectiveness of the provided approach was evaluated using the subsequent quality indicators:

1) ACCURACY - The caability of the classifier to accurately predict the class of labels is known as accuracy. It is calculable as follows:

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn}$$

(1)

2) SENSITIVITY - The criteria that are most frequently used in epidemiological and medical research are sensitivity and specificity, however the majority of statisticians in mathematical domains are not familiar with them. It evaluates the classifier's aptitude for making accurate assumptions about the positive class. The following formula is used to calculate sensitivity.

$$Sensitivity = \frac{Tp}{Tp+Fn}$$

(2)

3) SPECIFICITY - Specificity measures how well the classifier can predict the negative class. The following definition of the word "specificity"

$$Specificity = \frac{Tn}{Tn+Fp}$$

(3)

4) F-SCORE - The F-score assess statistics exams. Prediction accuracy is determined by F-score using Recall and Precision. It is also possible to determine the F-score by weighing recall and accuracy. The recall is calculated by dividing the total number of forecasts by the number of correct guesses. Precision is calculated by dividing the total number of forecasts by the number of forecasts that were accurate. The value of the F-score is determined by,

$$F - Score = 2 * \frac{Precision*Recall}{Precision+Recall}$$

(4)

Where $Precision = \frac{Tp}{Tp+Fp}$ and $Recall = \frac{Tp}{Tp+Fn}$

5) CONFUSION MATRIX - The truth and fiction of the machine learning method are depicted in a confusion matrix. The size of the confusion matrix is inversely proportional to the number of items that must be anticipated. The rows of the confusion matrix show the machine learning algorithm's forecast, and the columns of the confusion matrix show the actual value of the known truth. As a result, the top left and bottom right corners of Figure 7 each contain a genuine positive and a genuine negative. The False Positive is located in the top right-hand corner of Fig. 8, while the False Negative is in the bottom left-hand corner.

	P	N
P	True Positive (TP)	False Negative (FN)
T	False Positive (FP)	True Negative (TN)

Actual class appears to the left of the matrix.

Fig 7: Confusion Matrix

5.2 Result Analysis

The Table 2 shows the performance analysis of the proposed learners based ensemble system. It has resulted from Table 1 that VGG-16, VGG-19, and inception V3 provide the accuracy values of 92%, 87%, and 78% respectively. Table 1 also shows that the sensitivity values of VGG-16, VGG-19, and inception V3 are 86%, 80% and 69%, respectively. By considering mentioned values, the proposed VGG-16 ensemble model classifies the cancerous images even more accurately in comparison to the VGG-19 and inception V3. However, it is discovered that the specificity values of VGG-16, VGG-19, and inception V3 are 90%, 74% and 62%, respectively. While at the same time, it is noticed from table that the proposed ensemble model has a higher F- Score, lower False-positive, and higher precision values as compared to VGG-19 and inception V3. To add further, It can be notify from the table that the proposed ensemble performs better than the ensemble approach developed in [27] in terms of accuracy, sensitivity, and specificity.

Table 2: Comparison of performance between VGG-16, VGG-19 and inception V3

| Models	Accuracy	Sensitivity	Specificity	F-Score	Precision
VGG-16	92%	0.86	0.90	0.91	0.96
VGG-19	87%	0.80	0.74	0.85	0.86
Inception V3	78%	0.69	0.62	0.79	0.76

Fig. 8 the test and training accuracies and loss of the individual learners. The accuracy of the VGG-16 model is shown in Fig. 8(A). It is observed from the figure that the training Accuracy of VGG-16 is up to 98% and test accuracy 92%. The accuracy of the VGG-19 model is shown in Fig. 8 A. It is observed from the figure that the training accuracy of the VGG-19 model is 89% and test accuracy is 87%. Fig. 8 A shows that the training accuracy of the Inception V3 model is 88% and test accuracy is 78%. Fig 8 part B shows the loss plot of training and validation of VGG-16, VGG-19 and inception V3.

## A. Trainning and Validation Accuracy Plot
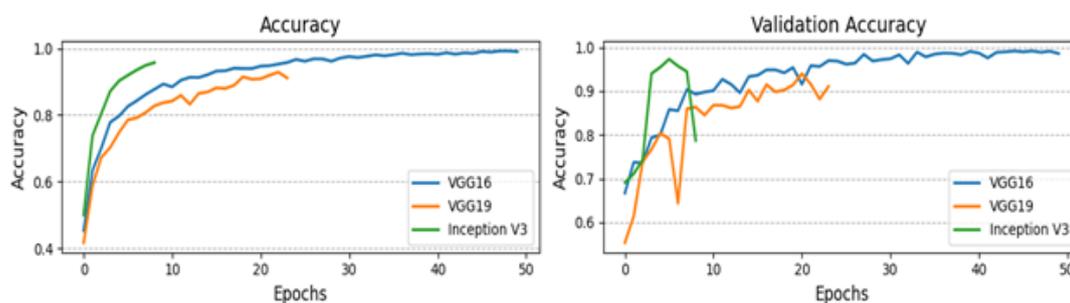
B. Trainning and Validation Loss Plot
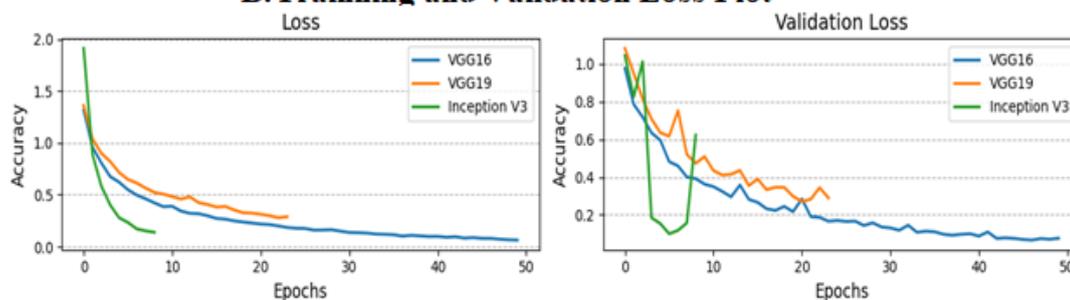
Fig 8. Training and Validation Accuracy and Loss plots

The confusion matrix is a crucial tool for assessing the performance of a classification model that aims to distinguish between cancer and non-cancer cases. In this scenario, VGG-16 model achieved an overall accuracy of 92%, which is an impressive result. Breaking down the confusion matrix, we find that there are 250 cancer cases (actual positives) and 320 non-cancer cases (actual negatives) in the dataset. Out of the 250 actual cancer cases, the model correctly predicted 230 of them (True Positives, TP), which reflects the model's ability to accurately identify cancer cases. However, it also incorrectly classified 20 actual cancer cases as non-cancer (False Negatives, FN). This indicates instances where the

model missed identifying cancer. On the non-cancer side, out of the 320 actual non-cancer cases, the model correctly predicted 295 as non-cancer (True Negatives, TN), showcasing its proficiency in accurately recognizing non-cancer cases. Unfortunately, it also made 25 incorrect predictions, classifying non-cancer cases as cancer (False Positives, FP) as shown in figure 9. This confusion matrix enables a more nuanced assessment of the model's performance. The 92% accuracy figure indicates the overall proportion of correct predictions, but the individual TP, TN, FP, and FN values offer insights into the model's strengths and weaknesses.
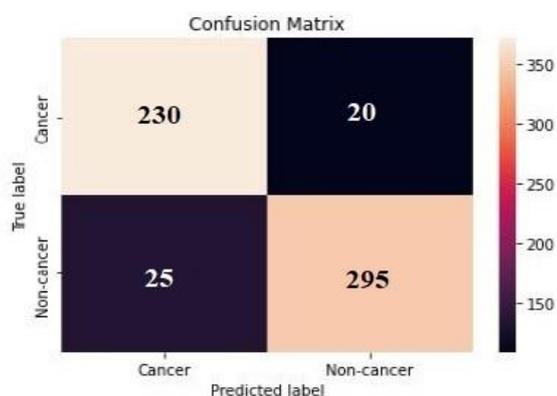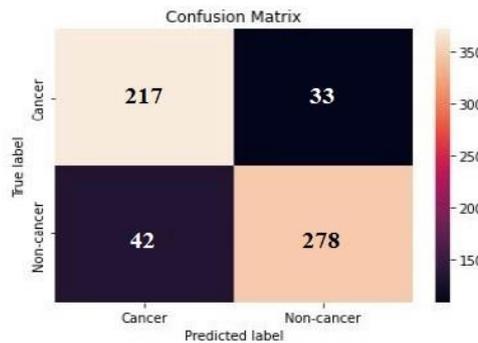

Fig 9. VGG-16 Confusion Matrix

The confusion matrix for VGG-19 model with 87% accuracy, using 250 cancer data and 320 non-cancer data points is as shown in figure 10. The confusion matrix

offers a detailed assessment of VGG-19 model's performance in classifying cancer and non-cancer cases. With an accuracy of 87%, the model has achieved a

commendable overall performance. However, to delve deeper into its strengths and limitations, the confusion matrix provides a more nuanced perspective. Out of the 250 actual cancer cases (actual positives), the model correctly identified 217 of them (True Positives, TP). This reflects the model's ability to accurately detect cancer cases. Conversely, it also misclassified 33 actual cancer cases as non-cancer (False Negatives, FN), highlighting instances where the model missed identifying cancer. On the non-cancer side, out of the 320 actual non-cancer cases (actual negatives), the model correctly predicted 278 as non-cancer (True Negatives, TN), showcasing its

proficiency in recognizing non-cancer cases. However, it also made 42 incorrect predictions, classifying non-cancer cases as cancer (False Positives, FP). These individual TP, TN, FP, and FN values enable a more detailed evaluation of the model's performance. Metrics like sensitivity, specificity, precision, and the F1-Score can provide a comprehensive understanding of its ability to correctly classify cancer and non-cancer cases. This insight is essential for fine-tuning the model and improving its performance, especially in critical healthcare applications like cancer detection.

Fig 10. VGG-19 Confusion Matrix

The confusion matrix for an Inception V3 model with 78% accuracy, using 250 cancer data and 320 non-cancer data points is shown in Figure 11. The confusion matrix for this Inception V3 model provides an insightful perspective on its classification performance, specifically in the context of distinguishing between cancer and non-cancer cases. While the model attains a 78% overall accuracy, the matrix unveils a more nuanced understanding of its capabilities. With 250 actual cancer cases (true positives) in the dataset, the model correctly identifies 195 of them (True Positives, TP), reflecting its competence in accurately recognizing cancer cases. However, there are instances where the model misses cancer cases, resulting in 55 false negatives (False Negatives, FN). On the non-cancer side, out of 320 actual

non-cancer cases, the model effectively classifies 249 as non-cancer (True Negatives, TN), illustrating its proficiency in identifying non-cancer cases. Nonetheless, the model also commits 71 incorrect predictions, labeling non-cancer cases as cancer (False Positives, FP). These individual TP, TN, FP, and FN values enable a more thorough evaluation of the Inception V3 model's performance. Metrics such as sensitivity, specificity, precision, and the F1-Score provide a comprehensive overview of its ability to correctly distinguish between cancer and non-cancer cases. In the domain of healthcare applications, particularly cancer detection, this detailed assessment is critical for ensuring the model's accuracy, reliability, and potential for saving lives through early diagnosis and intervention.
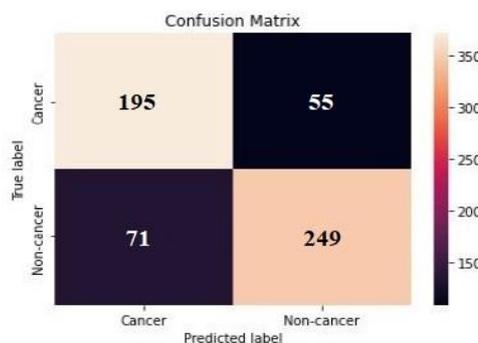
Fig 11. Inception V3 Confusion Matrix

The graphical illustration of various models, including VGG16, VGG19, and Inception V3, reveals an interesting trade-off between accuracy and training time, with VGG-16 standing out as a particularly weighted option due to its notable performance difference.

A. VGG16: VGG16, part of the VGGNet architecture family, is celebrated for its simplicity and remarkable accuracy in image classification tasks. The model consists of 16 weight layers, hence the name. Its design emphasizes stacking small-sized 3x3 convolutional kernels with max-pooling layers and ReLU activations, creating a deep and powerful network. Due to its relatively shallower structure compared to VGG-19, it generally offers faster training times. However, this slight trade-off in terms of the model's depth results in a subtle accuracy difference.

B. VGG19: VGG19, an extension of VGG16, takes the depth concept even further, employing 19 weight layers. This architectural evolution enhances feature learning capabilities, making it more proficient in capturing intricate image details. However, this comes at the cost of longer training times, given the additional layers. The trade-off here is a slight boost in accuracy due to its deeper architecture.

C. Inception V3: Inception V3, designed by Google, presents an intriguing alternative in this trade-off scenario. It follows a different architectural philosophy, with multiple branches and an efficient use of various filter sizes. This design allows it to capture multi-scale features effectively, resulting in competitive accuracy. Training times for Inception V3 are typically between VGG16 and VGG19, striking a balance between model depth and computational efficiency.

Despite the accuracy and training time trade-offs, VGG-16 emerges as the heavyweight in this comparison due to its significant performance difference. It offers impressive accuracy while requiring less training time than VGG-19 or Inception V3, making it a preferred choice for real-time or resource-constrained applications where speed is essential. On the other hand, VGG-19 and Inception V3, with their deeper architectures, shine when the highest levels of accuracy are required, and training time is not a critical factor. The choice among these models ultimately hinges on the specific task, available resources, and the desired balance between speed and accuracy.

6. Conclusion

The most common reason for skin cancer-related deaths is a malignant lesion. It could be treatable if discovered in its early stages. Deep learning techniques have been used for cancer detection in the literature; however, the effectiveness of the individual learners is constrained. For decision-making on delicate subjects like cancer, ensemble models can improve performance. An ensemble model to identify skin cancer was created in this article. It is demonstrated how vgg-16, vgg-19, and inception v3 are compared. The findings show that the suggested ensemble, VGG-16, has an average accuracy of 92%. Compared to VGG-19 and Inception V3, the suggested VGG-16 model performs better in terms of sensitivity, accuracy, F-Score, specificity, false-positive rate, and precision. We plan to investigate the effectiveness of reinforcement learning-based methods for skin cancer diagnosis in the future.

References

[1] A. Afroz, R. Zia, A. O. Garcia, M. U. Khan, U. Jilani, and K. M. Ahmed, "Skin lesion classification using machine learning approach: A survey," 2022. doi: 10.1109/GCWOT53057.2022.9772915.

[2] M. R. Grever, S. A. Schepartz, and B. A. Chabner, "The National Cancer Institute: Cancer drug discovery and development program," *Seminars in Oncology*, vol. 19, no. 6, 1992.

[3] L. E. Davis, S. C. Shalin, and A. J. Tackett, "Current state of melanoma diagnosis and treatment," *Cancer Biology and Therapy*, vol. 20, no. 11. 2019. doi: 10.1080/15384047.2019.1640032.

[4] U. Yadav and A. K. Sharma, "A novel automated depression detection technique using text transcript BGRU, deep learning, depression detection, depression levels, PHQ score, text embeddings, text transcripts," *Int J Imaging Syst Technol*, 2022, doi: 10.1002/ima.22793.

[5] T. Nyíri and A. Kiss, "Novel ensembling methods for dermatological image classification," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11324 LNCS. doi: 10.1007/978-3-030-04070-3_34.

[6] A. C. F. Gouabou, J. L. Damoiseaux, J. Monnier, R. Iguernaissi, A. Moudafi, and D. Merad, "Ensemble method of convolutional neural networks with directed acyclic graph using dermoscopic images: Melanoma detection application," *Sensors*, vol. 21, no. 12, 2021, doi: 10.3390/s21123999.

[7] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *Journal of Biomedical Informatics*, vol. 86, 2018, doi: 10.1016/j.jbi.2018.08.006.

[8] U. Yadav and A. K. Sharma, "Review on Automated Depression Detection from audio visual clue using Sentiment Analysis," in *Proceedings of the 2nd*

International Conference on Electronics and Sustainable Communication Systems, ICESC 2021, 2021, pp. 1462–1467. doi: 10.1109/ICESC51422.2021.9532751.

[9] Z. Waheed, A. Waheed, M. Zafar, and F. Riaz, "An efficient machine learning approach for the detection of melanoma using dermoscopic images," 2017. doi: 10.1109/C-CODE.2017.7918949.

[10] M. Sattar and A. Majid, "Lung Cancer Classification Models Using Discriminant Information of Mutated Genes in Protein Amino Acids Sequences," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, 2019, doi: 10.1007/s13369-018-3468-8.

[11] A. Majid, S. Ali, M. Iqbal, and N. Kausar, "Prediction of human breast and colon cancers from imbalanced data using nearest neighbor and support vector machines," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 3, 2014, doi: 10.1016/j.cmpb.2014.01.001.

[12] S. V. Khedikar and U. Yadav, "Detection of disease from radiology," in *Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017*, 2018, vol. 2018-Janua, pp. 1–4. doi: 10.1109/ICIIECS.2017.8276174.

[13] R. D. Seeja and A. Suresh, "Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM)," *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 5, 2019, doi: 10.31557/APJCP.2019.20.5.1555.

[14] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, 2017, doi: 10.1038/nature21056.

[15] H. A. Haenssle *et al.*, "Man against Machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, no. 8, 2018, doi: 10.1093/annonc/mdy166.

[16] T. T. Q. Trinh, Y. C. Chung, and R. J. Kuo, "A domain adaptation approach for resume classification using graph attention networks and natural language processing," *Knowledge-Based Systems*, vol. 266, 2023, doi: 10.1016/j.knosys.2023.110364.

[17] *Technology in Practical Dermatology*. 2020. doi: 10.1007/978-3-030-45351-0.

[18] M. Rajadhyaksha, S. González, J. M. Zavislan, R. R. Anderson, and R. H. Webb, "In vivo confocal scanning laser microscopy of human skin II: Advances in instrumentation and comparison with histology," *Journal of Investigative Dermatology*, vol. 113, no. 3, 1999, doi: 10.1046/j.1523-1747.1999.00690.x.

[19] "SkinVision | Skin Cancer Melanoma Detection App | SkinVision." https://www.skinvision.com/ (accessed Oct. 16, 2023).

[20] A. Aboulmira, E. M. Raouhi, H. Hrimech, and M. Lachgar, "Ensemble learning methods for deep learning: Application to skin lesions classification," 2022. doi: 10.1109/ISIVC54825.2022.9800732.

[21] Z. Lan, S. Cai, X. He, and X. Wen, "FixCaps: An Improved Capsules Network for Diagnosis of Skin Cancer," *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3181225.

[22] W. Cao *et al.*, "ICL-Net: Global and Local Inter-Pixel Correlations Learning Network for Skin Lesion Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 1, 2023, doi: 10.1109/JBHI.2022.3162342.

[23] A. Javaid, M. Sadiq, and F. Akram, "Skin Cancer Classification Using Image Processing and Machine Learning," 2021. doi: 10.1109/IBCAST51254.2021.9393198.

[24] A. Murugan, S. A. H. Nair, A. A. P. Preethi, and K. P. S. Kumar, "Diagnosis of skin cancer using machine learning techniques," *Microprocessors and Microsystems*, vol. 81, 2021, doi: 10.1016/j.micpro.2020.103727.

[25] W. Zhang, L. Gao, and R. Liu, "Using Deep Learning Method for Classification: A Proposed Algorithm for the ISIC 2017 Skin Lesion Classification Challenge," *arXiv [csCV]*, 2017.

[26] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3003890.

[27] A. Younis, L. Qiang, C. O. Nyatega, M. J. Adamu, and H. B. Kawuwa, "Brain Tumor Analysis Using Deep Learning and VGG-16 Ensembling Learning Approaches," *Applied Sciences (Switzerland)*, vol. 12, no. 14, 2022, doi: 10.3390/app12147282.

[28] Z. Cao, J. Huang, X. He, and Z. Zong, "BND-VGG-19: A deep learning algorithm for COVID-19 identification utilizing X-ray images," *Knowledge-Based Systems*, vol. 258, 2022, doi: 10.1016/j.knosys.2022.110040.

[29] J. Cao, M. Yan, Y. Jia, X. Tian, and Z. Zhang, "Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals," *Eurasip Journal on Advances in Signal Processing*, vol. 2021, no. 1, 2021, doi: 10.1186/s13634-021-00740-8.