Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

Skin Lesion Classification and Cancer Detection Using Deep Learning

[1]B.Meena*, [2]A.Keerthika , [3]B.Madhu Kiran , [4]Dileep Panigrahi

[1]meenacserec@gmail.com , [2]19981a0513@raghuenggcollege.in

[3]19981a0518@raghuenggcollege.in , [4]19981a0540@raghuenggcollege.in

Department of Computer Science Engineering

Raghu Engineering College -Visakhapatnam-531162

Abstract--*Skin cancer is regarded as one of the most perilous forms of cancer, and a significant rise in the number of fatalities has occurred due to insufficient awareness about the symptoms of this cancer and the necessary preventative measures.The generation of genetic defects or mutations on the skin due to unresolved DNA is typically the cause of skin cancer. Skin lesion classification and cancer detection have proven to be challenging for clinicians, owing to the variations in skin lesions and the absence of objective criteria. This study presents a deep learning methodology for classifying skin lesions and detecting cancer using Convolutional Neural Networks (CNNs).The dataset used is containing skin lesion images, and is consisting of different types of skin lesions, including cancerous and non-cancerous lesions. Our project includes a preprocessing method, where images are normalized, and lesions are segmented. For the classification of skin lesion types, Max Pooling 2D, Batch Normalization, Convo2D, Dense, and Dropout layers was implemented and the accuracy obtained is about 98.63%. The segmented lesions are then fed into a CNN model, which is trained to classify the lesions into cancerous or non-cancerous. The proposed approach is focusing on achieving high accuracy in skin lesion classification and cancer detection.*

Keywords: Skin Lesion Classification, CancerDetection, DeepLearning(DL), Convolution Neural Networks(CNN).

1. **Introduction**:

Skin cancer is a serious health problem worldwide, and its early detection is essential for successful treatment. Melanoma is responsible for 75% of deaths that occur due to skin cancer.A number of factors contribute to the development of skin cancer, including air pollution, UV radiation, unhealthy living habits, etc[2]. Dermatologists and healthcare professionals rely on visual inspection and biopsy for skin cancer diagnosis, which can be time-consuming and subjective. The accurate and timely diagnosis of skin cancer is critical to reducing the morbidity and mortality rates associated with this disease. Over the past few years, algorithms based on deep learning techniques have shown promising results in the automated classification of skin lesions and the detection of cancer, offering faster and more accurate results which are alternative to traditional methods. The use of deep learning algorithms for skin lesion classification and cancer detection has the potential to improvise the field of dermatology. By providing a reliable and

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

efficient method for automated diagnosis, this technology can help to reduce the workload of dermatologists, increase the accuracy of skin cancer detection, and ultimately improve patient outcomes. The motivation for this study is to explore the capability of deep learning algorithms for skin lesion classification and cancer detection and to investigate the performance of different architectures and techniques in this domain. The study aims to address the challenges associated with traditional methods of skin cancer diagnosis and to provide a more accurate, efficient, and accessible solution. Skin lesion classification and cancer detection using deep learning involve the use of machine learning algorithms to automatically classify dermatological images into different categories such as malignant, benign, or normal. The primary inquiries that are being explored in this investigation are:

1. What is the performance of deep learning algorithms for skin lesion classification and cancer detection?
2. How can different architectures and techniques be optimized for this task?
3. What are the benefits and limitations of using deep learning for skin cancer diagnosis?

Previous studies have investigated the use of various machine-learning approaches for skin lesion classification and cancer detection. These include support vector machines (SVMs), and random forests. Despite giving the promising results reported in previous studies, Further advancements are required in developing deep learning-based techniques that are both precise and effective in classifying skin lesions and detecting cancer. In addition, there is a need for more comprehensive evaluations of these approaches using larger datasets and different evaluation metrics like accuracy, precision, and recall. The problem addressed in this study is the accurate and appropriate classification of skin lesions and the detection of cancer using deep learning techniques. The proposed solution is to develop a deep learning-based approach that uses a CNN architecture for extracting features from dermatological images and categorizing them into distinct classes such as cancerous or non-cancerous. The system will be trained and evaluated using a publicly available dataset, and its performance and accuracy will be compared to methods currently available for classifying skin lesions and detecting cancer. The purpose of this study is to develop and evaluate a deep learning-based approach for skin lesion classification and cancer detection. The contribution of this study is the development and evaluation of a an approach based on deep learning techniques for classifying skin lesions and detecting cancer.

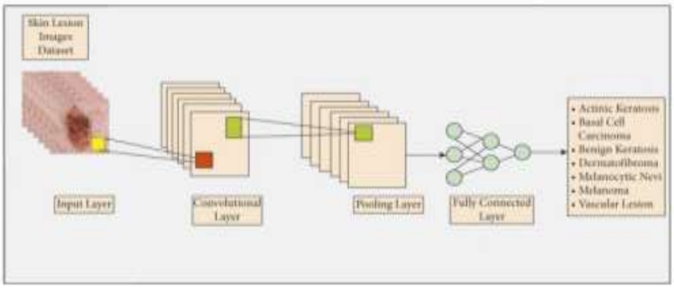


Fig.1

Fig.1 illustrates the architecture of the CNN sequential model proposed in this study.

2. Literature Review(LR):

Skin cancer poses a significant public health issue, with over one million new cases being diagnosed annually worldwide. Early detection and accurate diagnosis of skin lesions can significantly increase the chances of successful treatment. Techniques in machine learning, such as support vector machines (SVM), decision trees, and random forests, have shown great promise in the classification and detection of skin lesions.Several studies have investigated the use of machine learning for skin lesion classification. In the research made by Codella et al. [3](2018), a multi-class SVM was trained on a dataset of over 10,000 clinical images, achieving an accuracy of 75.5%. Similarly, Haenssle et al.[4] (2018) trained a decision tree on a dataset of over 2,000 clinical images, achieving an accuracy of 86.8%. Skin cancer represents a significant public health challenge as it is diagnosed in more than one

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

million new cases every year on a global scale. Early detection and accurate diagnosis of skin lesions can significantly increase the chances of successful treatment. The use of machine learning techniques such as support vector machines (SVM), decision trees, and random forests has shown considerable potential in detecting and classifying skin lesions.Other studies have focused in the context of skin cancer detection using machine learning. In a study by Gutman et al.[5] (2016), a random forest was trained on a dataset of over 2,000 clinical images, achieving a sensitivity of 92.1% and a specificity of 86.1%.

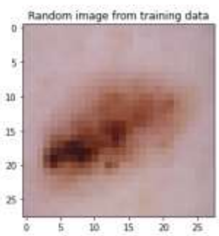
3. Methodology:

Skin cancer is a major health concern worldwide, with millions of new cases diagnosed each year. Early detection and accurate diagnosis are critical for successful treatment and improved patient outcomes. Over the past few years, convolutional neural networks (CNNs), which are a type of deep learning technique, have emerged as promising tools for skin cancer detection and classification., pooling the results, and passing the output through non-linear activation functions. This allows the network to learn hierarchical representations of the input image, with lower-level features detected in early layers and more complex features detected in later layers. When it comes to detecting skin cancer, CNNs have the capability to undergo training on extensive datasets of skin Convolutional Neural Networks (CNNs) are a neural network architecture that has demonstrated remarkable performance in image recognition tasks. The network is composed of several layers of neurons that learn to identify patterns in images by applying filters to the input image,lesion images, whereby labels indicate whether the lesions are benign or malignant.The network learns to recognize features that are associated with malignancy, such as asymmetry, irregular borders, and color variation. Once trained, the CNN can be used to classify new images of skin lesions as either malignant or benign, providing a powerful tool for early detection and diagnosis. When designing a research study on skin lesion classification and cancer detection using CNN, several factors need to be considered, including the choice of the dataset, CNN architecture, evaluation metrics, and experimental design.

Dataset Selection: Choose a publicly available dataset of skin lesion images, such as the ISIC Archive or the HAM10000 dataset.Data imbalances and data limitations in skin-disease datasets were addressed by the International Skin Imaging Collaboration (ISIC)[6]. Ensure that the dataset includes a sufficient number of malignant and benign lesions and that the images are labeled accurately.

Data Preprocessing: Resize the images in the dataset to a uniform size as a preprocessing step, normalizing the pixel values, and dividing the data into sets for training and testing purposes and Enhancing images is about intensifying their quality[7]. various data preprocessing techniques include

Fig.2



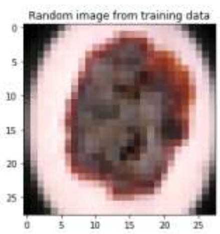


Fig.3 Fig.4

From fig.2,fig.3,fig.4 we can identify that random images are taken while the data pre-processing step is in the process.

(i) Image resizing: Resizing images to a uniform size helps to reduce the computational complexity of the model and makes it easier to handle images of different resolutions. This can be done using techniques such as bilinear interpolation, bicubic interpolation, or nearest neighbor interpolation.

(ii) Image normalization: Normalizing images involves transforming the pixel values of the adjusting and scaling the input of a layer to have an average of zero and a standard deviation of one. This helps to improve the convergence of the model during training and also reduces the effect of illumination changes on the image.

(iii) Image augmentation: Image augmentation involves creating new training images Transformations can be applied to the original images, such as rotations, translations, flips, as well as adjustments to brightness and contrast. Augmenting images can enhance the variety of training data, ultimately boosting the model's capacity to generalize to novel, unobserved data.

(iv) Color normalization: Color normalization involves removing color variations in images that are caused by differences in illumination or camera settings. This can be done using techniques such as Macenko's method, Reinhard's method, or Vahadane's method.

(v) Noise reduction: Image noise can be caused by factors such as sensor noise, compression artifacts, or motion blur. Noise reduction techniques such as median filtering or Gaussian filtering can be applied to remove noise and improve the clarity of the image.

(vi) Edge detection: Edge detection algorithms such as Canny edge detection or Sobel edge detection can be used to highlight the edges in an image. This can be useful in applications such as object detection, where edges can provide important cues about the location and shape of objects in the image.

CNN Architecture: Design a CNN architecture with Conv2D, max pooling 2D, batch normalization, flattening, and dense, dropout layers.An object's features are extracted by CNN[8]. Consider incorporating techniques such as dropout and batch Normalization to prevent overfitting.

(i) Dropout: Dropout is a regularization technique used to prevent overfitting in deep learning models. It randomly drops out a certain fraction of the input units during training, forcing the model to learn more robust features.

The formula for dropout is:

Dropout = (Input * Mask) / (1 - Dropout Rate)

where Input is the input tensor, Mask is a binary mask generated from a Bernoulli distribution with probability (1 - dropout rate), and Dropout Rate is the probability of dropping out an input unit.

(ii)Flatten: The Flatten layer is utilized to transform the output of a convolutional layer into a one-dimensional vector. This is typically done before feeding the output to a fully connected layer.

The formula for flattening is simple:

Flatten = Reshape(-1)

where Reshape(-1) flattens the input tensor into a 1-dimensional vector.

(iii)Conv2D: Conv2D is a layer that performs 2-dimensional convolution on the input tensor. It employs a collection of filters to extract spatial characteristics from the input.
The formula for Conv2D is:

Conv2D = Convolution(Input, Filters) + Bias

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

(iv)Dense Layer: Dense layers, also known as fully connected layers, are the traditional layers used in artificial neural networks. They are used to map the output of the previous layer to a set of desired output classes. Dense layers take in a flattened input and output a one-dimensional array of values.

The formula for a dense layer is:

Output=ativation_function(dot(weights) + bias)

(v)Batch Normalization Layer: Batch Normalization is a normalization technique that

adjusts and scales the input of a layer to attain zero mean and unit variance. By minimizing the internal covariate shift, this technique enhances the stability and performance of the model.

The formula for a batch normalization layer is:

normalized_output = (input - mean) / sqrt(variance + epsilon)
output = gamma * normalized_output + beta

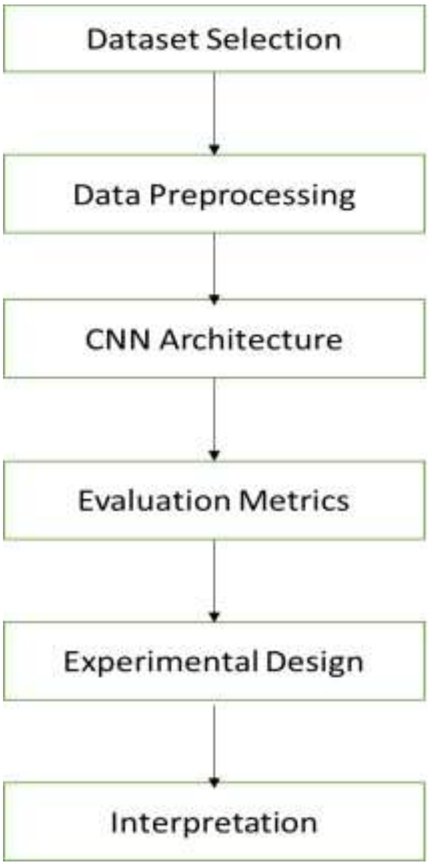


Fig.5

The fig.5 gives the flow of the proposed approach

Evaluation Metrics: Choose appropriate evaluation metrics, such as accuracy, precision, recall, and F1 score, to measure the performance of the CNN. Additionally, use techniques such as ROC curves and confusion matrices to visualize the performance of the CNN.

i. Epoch: The number of epochs is typically determined by the user or the data scientist and is set before the training process begins. It represents the number of times the model will iterate through the entire dataset during training. For example, if a model is trained for 10 epochs on a dataset with 1000 samples, the model will see all 1000 samples 10 times during training.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

ii. ETA: ETA is an estimate of the time remaining for the current epoch or the entire training process. It is calculated based on the time taken to complete the previous epochs and the remaining epochs.

The formula for calculating ETA can be expressed as follows:

ETA = (total_epochs - current_epoch) * (time_taken_for_previous_epochs/ current_epoch)

iii. Accuracy: Accuracy is a commonly used evaluation metric in machine learning that measures the percentage of correct predictions made by a model.

The formula for calculating accuracy is:

Accuracy = (NP / TP) x 100%

NP=Number of correct predictions.

TP=Total number of predictions.

Experimental Design: Train the CNN on the training data using the chosen architecture and hyperparameters, and evaluate its performance on the testing data. Use matplotlib library to estimate the model's performance on unseen data, and perform ablation studies to analyze the impact of individual layers or hyperparameters on the model's performance.

Interpretation: Interpret the results of the study and conclude the effectiveness of CNN for skin lesion classification and cancer detection. Additionally, discuss potential limitations of the study, such as dataset bias or overfitting, and propose future research directions to address these limitations.

The analysis made out of our project is based on various factors like

(i)High Accuracy: CNNs have shown high accuracy rates in classifying skin lesions and detecting skin cancer. In many studies, CNNs have achieved accuracy rates of over 90%.

(ii)Non-invasive: Skin lesion classification and cancer detection using CNNs are non-invasive methods that can be easily implemented in a clinical setting. This can help to reduce the number of unnecessary biopsies and improve patient outcomes.

(iii)Speed: CNNs can analyze skin lesions and detect skin cancer much faster than human experts. This can help to improve the efficiency of diagnosis and treatment.

There are various limitations of the proposed system which classified as

i. Limited Data: CNNs require a large amount of data to be trained effectively. However, there is a limited amount of high-quality data available for skin lesion classification and cancer detection. This can lead to overfitting, where the model learns to classify the training data accurately but does not generalize well to new data.

ii. Bias: The accuracy of CNNs can be affected by bias in the training data. For example, if the training data contains more images of skin lesions from certain demographic groups, the model may be less accurate at classifying lesions from other groups.

iii. False Positives and False Negatives: CNNs can produce false positives (expressing an instance of categorizing a non-cancerous lesion as cancerous.) and false negatives (Incorrectly identifying a malignant lesion as benign.). These errors can lead to unnecessary treatments or missed diagnoses.

iv. Interpretability: CNNs are often seen as a black box, meaning that it can be difficult to

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

understand how the model makes its decisions. This can be a limitation for doctors who need to explain the reasoning behind the model's classifications to their patients.

4. Results:

EPOCH=10	ETA (Seconds and ms)	ACCURACY
1	8s 24ms/step	0.5744
2	5s 20ms/step	0.8618
3	4s 19ms/step	0.9108
4	4s 19ms/step	0.9324
5	5s 20ms/step	0.9473
6	4s 19ms/step	0.9538
7	4s 19ms/step	0.9601
8	5s 19ms/step	0.9653
9	5s 20ms/step	0.9653
10	5s 19ms/step	0.9724

Fig.6

The above fig.6 (table) gives the result of how well the model fits into the training data. The metrics considered here are epoch which is a complete pass through an algorithm and ETA value and followed by accuracy. The results are taken for a sample of 10 epochs which means the model will see all the 10000 samples 10 times and their accuracies are analyzed.

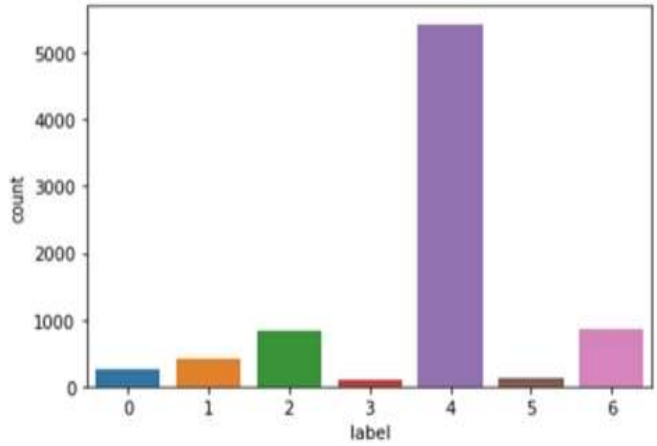


Fig.7

From fig.7 we can observe that cell type Melanocytic nevi has a very large number of instances in comparison to other cell types.

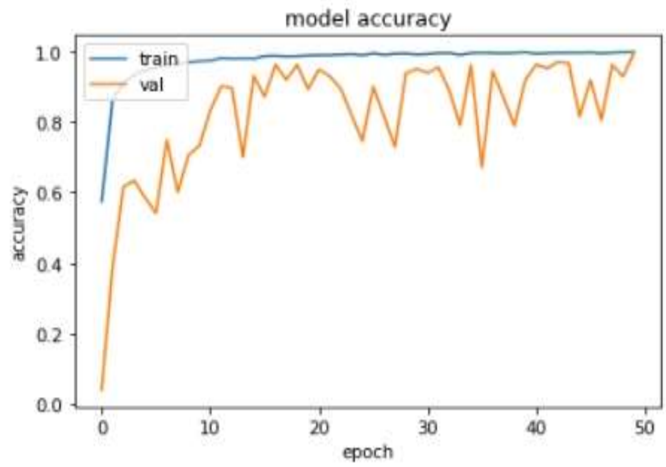


Fig.8

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

From fig.8 we can describe that Model accuracy is usually Precision is a performance metric for a model, calculated as the ratio of the number of true positives (correctly predicted positive cases) to the total number of positive predictions made by the model. It's one way of assessing the performance of a model, but certainly, it's not the only way. According to our proposed approach, the below graph shows the model accuracy which is how well testing data is matching with the training data set classifications.

A set of test cases that are tested for the proposed approach include:

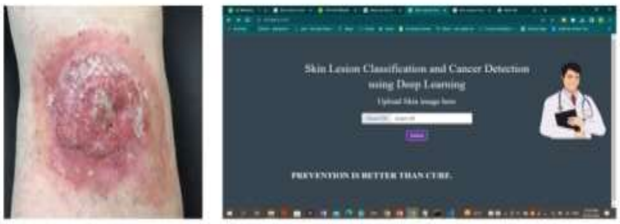


Fig.9

Fig.9 shows the image that is uploaded for the skin cancer classification as well as for detection.

Fig.10

Fig.10 is the corresponding output for fig.9 that is the image is identified as cancer and the lesion belongs to the classification called actinic keratoses and intraepithelial carcinoma.

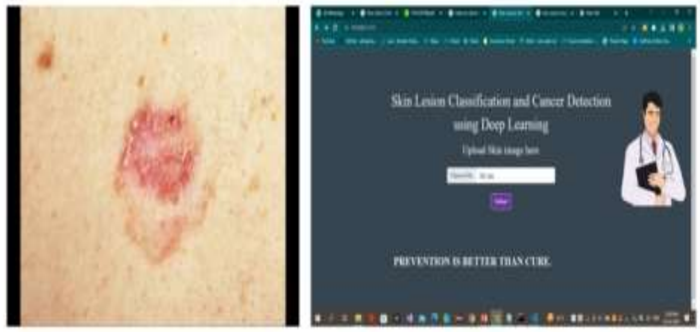


Fig.11

Fig.11 shows the image that is uploaded for the skin cancer classification as well as for detection.

Fig.12

Fig.12 is the corresponding output for fig.11 and here the image is identified as non-cancerous and the lesion belongs to the classification called dermatofibroma.

5. Discussion:

Skin lesion classification and cancer detection using CNNs (Convolutional Neural Networks) is an active and important research area in medical image analysis. The database website was consulted when studies indicated they used a specific database, but did not provide specific details of the included lesion types[9]. When it comes to classifying skin lesions and detecting cancer, CNNs can undergo training on extensive datasets of labeled images, allowing them to automatically learn pertinent features and patterns. The features can subsequently be employed to categorize new images as either benign or malignant or to detect particular skin lesion types.One of the key challenges in this field is the availability of high-quality annotated datasets, which are essential for training and evaluating machine learning algorithms.The application of pre-processing techniques, such as hair removal, shadow removal, glare removal, and segmentation, presents another challenge.[10].Despite these challenges, there have been several promising studies demonstrating the effectiveness of CNNs for skin lesion classification and cancer detection. For example, a study published in the journal Nature in 2018 showed that a CNN outperformed dermatologists in classifying skin lesions as benign or malignant. Another study published in the Journal of the American Academy of Dermatology in 2019 demonstrated that a CNN was able to accurately detect a specific type of skin lesion (seborrheic keratosis) with high sensitivity and specificity.

6. Conclusion:

In summary, skin lesion classification and cancer detection using CNNs is a promising approach to improving the accuracy and efficiency of dermatological diagnoses. CNNs have been shown to outperform dermatologists in some studies, and have demonstrated high sensitivity and specificity in detecting specific types of skin lesions. In order to achieve high accuracy and precision in skin cancer classification, the proposed system employs a CNN algorithm that utilizes a sequential model from Keras. This approach involves two effective layers, namely Dense and Batch Normalization, for feature extraction. The system uses seven types of skin cancers for classification, and it has achieved an accuracy of approximately 98.63%. In the future, enhancements to skin lesion classification and cancer detection using CNNs could include the development of more comprehensive datasets that include a wider range of skin lesions and populations, the use of explainable AI techniques to improve the interpretability and transparency of the algorithms, and the integration of other types of data such as genomic or histological information to further improve accuracy and precision.

7. References

[1] Ilker Ali OZKAN,Murat KOKLU,"Skin Lesion Classification using Machine Learning Algorithms",International journal of international systems and applications in engineering(IJISAE),Volume 05, Issue 03, April 2018.

[2]Praveen Banasode, Minal Patil and Nikhil Ammanagi,"A Melanoma Skin Cancer Detection Using Machine Learning Technique: Support Vector Machine",IOP conference Series :Materials science and engineering(ICFEST),doi:10.1088/1757-899X/1065/1/012039.

[3]kinnor das,Clay J.cockrell,Anant Patil,Pawel Pietkiewicz,Mario giulini,Stephan Grabbe,Mohamad goldust,"Machine Learning and Its Application in Skin Cancer",International journal of research and public health,PMID: 34949015,doi: 10.3390/ijerph182413409.

[4]Mohammad Ali Kadampur , Sulaiman AI Riyaee,"Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images",Informatics in Medicine Unlocked(IS),Volume 18,2020,100282.

[5]Michael A. Marchetti , Noel C.F. Codella , Stephen W. Dusza , David A. Gutman MD, Brian Helba d, Aadi Kalloo, Nabin Mishra, Cristina Carrera, M. Emre Celebi , Jennifer L. DeFazio, Natalia Jaimes, Ashfaq A. Marghoob, Elizabeth Quigley, Alon Scope, Oriol Yélamos, Allan.Halpern , "International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge",Journal of American academy and dermatology,Volume 78,Issue 2,February 2018.

[6]Yinhao Wu, Bin Chen, An Zeng, Dan Pan, Ruixuan Wang and Shen Zhao1*,"Skin Cancer Classification With Deep Learning: A Systematic Review",Frontiers in Oncology,Volume 12 - 2022 https://doi.org/10.3389/fonc.2022.893972,13 July 2022.

[7]M. K. Monika, N. Arun Vignesh, C. Usha Kumari et al., "Skin cancer detection and classification using machine learning, MaterialsTodayProceedings, https://doi.org/10.1016/j.matpr.2020.07.366.

[8]Christopher Ebuka Ojukwu,"Melanoma Skin Cancer Detection Using Support Vector Machines and Convolutional Neural Networks",International Journal of Scientific Research in Computer Science and Engineering,E-ISSN: 2320-7639,Vol.9, Issue.6, pp.09-12, December (2021) .

[9]O T Jones, R N Matin, M van der Schaar, K Prathivadi Bhayankaram, C K I Ranmuthu, M S Islam, D Behiyat, R Boscott, N Calanzani, J Emery,H C Williams, F M Walter,"Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review",Lancet Digit Health 2022,Volume 4 , Issue 6,E466-E476, JUNE 2022,DOI:https://doi.org/10.1016/S2589-7500(22)00023-1.

[10]Vijayalakshmi M M,"Melanoma Skin Cancer Detection using Image Processing and Machine Learning",International Journal of Trend in Scientific Research and Development(IJTSRD),ISSN: 2456-6470, Volume-3 ,Issue-4, June 2019, pp.780-784, URL: https://www.ijtsrd.com/papers/ijtsrd23936.pdf.