A Study on the Application of Machine Learning and Deep Learning Techniques for Skin Cancer Detection

Hritwik Ghosh, Irfan Sadiq Rahat, Sachi Nandan Mohanty, J. V. R. Ravindra, Abdus Sobur

Abstract-In the rapidly evolving landscape of medical diagnostics, the early detection and accurate classification of skin cancer remain paramount for effective treatment outcomes. This research delves into the transformative potential of artificial intelligence (AI), specifically deep learning (DL), as a tool for discerning and categorizing various skin conditions. Utilizing a diverse dataset of 3,000 images, representing nine distinct skin conditions, we confront the inherent challenge of class imbalance. This imbalance, where conditions like melanomas are over-represented, is addressed by incorporating class weights during the model training phase, ensuring an equitable representation of all conditions in the learning process. Our approach presents a hybrid model, amalgamating the strengths of two renowned convolutional neural networks (CNNs), VGG16 and ResNet50. These networks, pre-trained on the ImageNet dataset, are adept at extracting intricate features from images. By synergizing these models, our research aims to capture a holistic set of features, thereby bolstering classification performance. Preliminary findings underscore the hybrid model's superiority over individual models, showcasing its prowess in feature extraction and classification. Moreover, the research emphasizes the significance of rigorous data pre-processing, including image resizing, color normalization, and segmentation, in ensuring data quality and model reliability. In essence, this study illuminates the promising role of AI and DL in revolutionizing skin cancer diagnostics, offering insights into its potential applications in broader medical domains.

Keywords—Artificial intelligence, machine learning, deep learning, skin cancer, dermatology, convolutional neural networks, image classification, computer vision, healthcare technology, cancer detection, medical imaging.

I. INTRODUCTION

S KIN cancer, a common health concern globally, has been the subject of numerous research efforts due to its potential severity and high incidence rate. The cornerstone of successful treatment is the early detection and accurate classification of the various types of skin cancer. However, traditional diagnostic procedures, which rely heavily on visual examination and biopsy, can be invasive, time-consuming, and subject to human error. This underscores the need for diagnostic tools that are not only more accurate but also efficient. In this context, AI, particularly DL, has shown immense promise. DL, a subset of AI, employs algorithms that simulate the neural networks in the human brain, enabling them to 'learn' from vast quantities of data.

The objective of this study is to leverage the capabilities of DL for the classification of skin cancer. We utilized a dataset of 3000 images collected from the field, encompassing nine distinct skin conditions, including various types of carcinomas and melanomas. However, the dataset presents a challenge of class imbalance. This can skew the learning process towards the majority class. To counter this, we incorporated class weights in our model training, ensuring equal consideration for all classes, irrespective of their frequency. Our methodology involves the creation of a hybrid model that amalgamates the strengths of two pre-trained CNNs: VGG16 and ResNet50. These models, pre-trained on the ImageNet dataset, are capable of extracting intricate features from images. By integrating these models, we aim to capture a more comprehensive set of features, thereby enhancing the classification performance. The findings of this study hold significant potential for the field of dermatology, possibly leading to a more precise and efficient diagnosis of skin cancer. Moreover, our research adds to the growing body of knowledge in the realm of AI in healthcare, showcasing the potential of DL and hybrid models in the analysis of medical images.

II. LITERATURE SURVEY

Thomas [1] delved into DL methods for non-melanoma skin cancer characterization. The study emphasized the significance of aligning with laboratory and pathologist workflows. It explored four key areas: model interpretability, whole-tissue segmentation, generative modeling for image creation, and natural language processing for report generation. The goal was to create automated yet interpretable diagnostic systems.

Nawaz et al. [2] proposed an automated method for earlystage melanoma skin cancer segmentation using a DL approach, specifically faster region-based convolutional neural networks (RCNN), combined with fuzzy k-means clustering (FKM). The method preprocesses images to enhance visual information, then employs faster-RCNN for feature extraction, followed by FKM for melanoma segmentation. Tested on three standard datasets (ISBI-2016, ISIC-2017, and PH2), the approach achieved impressive accuracies, outperforming existing methods.

Khamparia et al. [3] introduced a DL framework driven by the Internet of Health Things (IoHT) for the detection and

Hritwik Ghosh is with School of Computer Science and Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India (e-mail: me.hritwikghosh@gmail.com).

Irfan Sadiq Rahat* and Sachi Nandan Mohanty are with SCOPE, VIT-AP University, Amaravati, Andhra Pradesh, India (*corresponding author, phone: 6300550795, e-mail: me.rahat2020@gmail.com, sachinandan09@gmail.com).

J. V. R. Ravindra is with Center for Advanced Computing Research Laboratory (C-ACRL), Department of Electronics and Communication Engineering, Vardhaman College of Engineering (Autonomous), Hyderabad, Telangana, India (e-mail: jayanthi@ieee.org).

Abdus Sobur is with Masters of Information Technology, Westcliff University, USA (e-mail: ronysobur@gmail.com)

classification of skin cancer using transfer learning. The framework utilizes pretrained architectures like VGG19, Inception V3, ResNet50, and SqueezeNet for feature extraction, which are then processed by a CNN for classifying skin cells. Integrated with the IoHT, this system aids medical specialists in diagnosing and treating skin cancer remotely. The proposed approach demonstrated superior performance in terms of precision, recall, and accuracy compared to other architectures.

Dildar et al. [4] conducted a systematic review on DL techniques for early detection of skin cancer. The study highlights the importance of early detection due to the dangerous nature of skin cancer and its potential to spread. The paper analyzes various techniques that utilize lesion parameters such as symmetry, color, size, and shape to differentiate between benign skin cancer and melanoma. The review consolidates findings from research papers published in reputable journals, presenting them through tools, graphs, tables, techniques, and frameworks for enhanced comprehension.

Quadir et al. [5] presented a study on skin cancer detection using CNN to classify skin lesions. The research focused on differentiating between benign moles and skin cancers, particularly highlighting melanoma's high risk. The study utilized a dataset of 10,000 clinical images, which underwent pre-processing and sharpening. The methodology involved constructing a neural network for accurate lesion edge detection and designing a mobile-friendly model. Using transfer learning, the research achieved high prediction accuracy and compared results with various models. The system effectively categorized different skin conditions, emphasizing the attributes extracted by the deep CNN.

Garcia [6] delved into the potential of meta-learning for skin cancer detection using DL techniques on dermoscopic images. The research aimed to understand the impact of leveraging knowledge from non-medical data on the classification of medical data, especially when faced with limited data. By finetuning a ResNet model pre-trained on non-medical data with a small sample from three different medical datasets, the study observed a 20-point performance increase in detecting melanoma and distinguishing between malignant and benign moles. The results underscored the value of non-medical image features in skin mole classification and highlighted the significance of data distribution on model performance.

Gomathi et al. [7] introduced a novel dual optimization based deep learning network (DODL net) for skin cancer detection. Utilizing the MNIST HAM10000 dataset, dermoscopic images underwent pre-processing with an adaptive median filter before segmentation in U-net. The hybridization of bacterial foraging optimization (BFO) and particle swarm optimization (PSO) was employed for feature extraction from segmented images. The deep CNN then classified seven different skin cancer types based on these features. The DODL net achieved an impressive accuracy of 98.76% on the MNIST HAM10000 dataset.

Sivakumar et al. [8] proposed a novel multidimensional Bregman divergence feature scaling based cophenetic piecewise regression recurrent deep learning classification (MBDFS-CPRRDLC) technique for early-stage skin cancer detection. Utilizing data from the medical internet of things (MIoT) stored in mobile cloud servers, the method employs multiple layers for feature selection and classification. The MBDFS technique identifies significant features, reducing time consumption, while the cophenetic correlative piecewise regression analyzes testing and training data. The proposed approach demonstrated increased accuracy and reduced cancer detection time compared to traditional methods.

Balambigai et al. [9] developed an optimized CNN model using random search optimization for classifying seven types of skin cancer. Using 10,015 images from the Human Against Machine dataset (HAM10000), the study addressed the challenge of selecting hyper-parameters for CNN models. While the base CNN model achieved an accuracy of 73.34%, the optimized model using random search optimization demonstrated an improved accuracy of 77.17%.

Fraiwan and Faouri [10] investigated the potential of raw deep transfer learning for classifying skin lesion images into seven categories. Emphasizing the importance of early diagnosis in skin cancer treatment, they developed a system using the HAM1000 dataset of dermoscopy images, employing 13 deep transfer learning models without explicit feature extraction or preprocessing. Despite achieving high accuracy for some cancer types, challenges like dataset imbalance and limited images in certain categories led to a maximum overall accuracy of 82.9%.

Pan et al. [11] highlighted the increasing risk of melanoma skin cancer due to UV ray exposure and presented an early detection mechanism for melanoma using image processing and DL. The image processing techniques, including threshold, edge detection, and geometry-based feature extraction, were employed to segment melanoma features based on the ABCDE criteria (asymmetry, border, color, diameter, and evolving). The DL model was then trained to predict the risk level of melanoma skin cancer. The proposed e-health application demonstrated high accuracy in melanoma detection.

Tembhurne et al. [12] emphasized the importance of accurate diagnosis of skin cancer, particularly melanoma, due to its fatal nature. They introduced a novel method that combines machine learning and DL techniques for skin cancer detection. The approach leverages state-of-the-art neural networks for feature extraction from images and machine learning models to process image features derived from techniques like contourlet transform and local binary pattern histogram. By integrating manual and automated features, their model achieved a 93% accuracy rate, with recall scores of 99.7% for benign and 86% for malignant cancers. Tested on the Kaggle dataset from the ISIC Archive, their ensemble method surpassed expert dermatologists and other advanced models, offering a valuable tool to reduce misdiagnoses.

Wang et al. [13] highlighted melanoma as the most lethal skin cancer, emphasizing the significance of early detection to increase survival rates. The study evaluated the performance of a pre-trained model, the Visual Geometry Group network (VGG), on the International Skin Imaging Collaboration (ISIC) 2019 challenge dataset for automatic melanoma classification. The model achieved a remarkable accuracy of 0.9067 and an AU ROC exceeding 0.93. Through ablation studies, the research identified various factors influencing the model's performance, such as training data size, frozen layers, classifier nodes, and data augmentation techniques.

Daghrir et al. [14] emphasized the critical nature of melanoma, a deadly skin cancer that can metastasize if not diagnosed early. The study introduced a hybrid approach for melanoma detection, leveraging both DL and classical machine learning techniques. The system utilized a CNN alongside two traditional machine learning classifiers, trained on features describing the borders, texture, and color of skin lesions. By combining the predictions of these three methods through majority voting, the approach achieved enhanced performance, with experiments indicating that the integrated method yielded the highest accuracy.

Singh et al. [15] addressed the global challenge of skin cancer detection, emphasizing the significance of early diagnosis for effective treatment. The study leveraged advancements in AI, particularly in healthcare, to detect skin cancer using machine learning and AI techniques. The research employed CNN for image processing and recognition, specifically utilizing models like VGG-16, MobileNet, and Inception-V3. The paper also provided an overview of various AI-based skin cancer detection models. Transfer learning was utilized to repurpose pre-trained models, and a unique model was also constructed from scratch using CNN blocks. Additionally, a web application was developed using HTML, Flask, and CSS.

Ghosh et al. [16] delved into the realm of water quality assessment using predictive machine learning. The study highlighted the efficacy of machine learning in evaluating and categorizing water quality. The dataset incorporated key parameters such as pH, dissolved oxygen, BOD, and TDS. Of the multiple models tested, the random forest model stood out, boasting an impressive accuracy of 78.96%. On the other hand, the SVM model trailed, achieving a lesser accuracy of 68.29%.

Ghosh et al. [17] explored the use of CNN models, specifically VGG19, DenseNet121, and ResNet50, for detecting potato leaf diseases. Using a comprehensive dataset of potato leaf images, they found VGG19 to be the most effective, followed closely by DenseNet121 and ResNet50. Their research highlights the potential of DL in aiding timely disease diagnosis, aiming to enhance potato crop health and productivity globally (Table I).

TABLE I	

	SUMMAR	Y OF THE LITERATURE SURVEY	
Reference	Focus of Study	Techniques Used	Key Findings
Thomas [1]	Non-melanoma skin cancer characterization	DL methods: model interpretability, whole- tissue segmentation, generative modeling, NLP	Automated yet interpretable diagnostic systems
Nawaz et al. [2]	Early-stage melanoma segmentation	DL: faster RCNN with FKM	High accuracies on standard datasets, superior to existing methods
Khamparia et al. [3]	Skin cancer detection and classification	IoHT, transfer learning with CNN	Superior performance in precision, recall, and accuracy
Dildar et al. [4]	Early detection of skin cancer techniques review	DL techniques analysis	Consolidated findings for early detection, importance of lesion parameters
Quadir et al. [5]	Skin cancer detection with CNN	CNN for lesion classification	High prediction accuracy, effective categorization
Garcia [6]	Meta-learning for skin cancer detection	ResNet with non-medical data	20-point performance increase in melanoma detection
Gomathi et al. [7]	DODL net for skin cancer detection	BFO and PSO with CNN	98.76% accuracy on MNIST HAM10000 dataset
Sivakumar et al. [8]	MBDFS-CPRRDLC technique for skin cancer detection	MBDFS with piecewise regression	Increased accuracy and reduced detection time
Balambigai et al. [9]	Optimized CNN model for skin cancer classification	Random search optimization for CNN	Improved accuracy of 77.17%
Fraiwan and Faouri [10]	Raw deep transfer learning for skin lesion classification	Deep transfer learning models	Max overall accuracy of 82.9%
Pan et al. [11]	Early detection of melanoma	Image processing and DL	High accuracy in melanoma detection
Tembhurne et al. [12]	Accurate diagnosis of skin cancer	ML and DL techniques	93% accuracy rate, outperformed expert dermatologists
Wang et al. [13]	Melanoma classification using VGG	VGG on ISIC 2019 dataset	Accuracy of 0.9067, AU ROC over 0.93
Daghrir et al. [14]	Melanoma detection hybrid approach	CNN with machine learning classifiers	-
Singh et al. [15]	Skin cancer detection using AI	CNN, VGG-16, MobileNet, Inception-V3	Development of web application, use of transfer learning
Ghosh et al. [16]	Water quality assessment with machine learning	Random forest, SVM	Random forest model accuracy of 78.96%
Ghosh et al. [17]	Potato leaf disease detection with CNN	CNN models: VGG19, DenseNet121, ResNet50	VGG19 found most effective

III. DATASET OVERVIEW

The dataset employed in our research is a comprehensive collection of 3000 images, meticulously gathered from the field. This dataset is unique and diverse, encompassing a wide range of skin conditions, specifically nine different classes. These classes include actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, and vascular lesion. Each class represents a distinct skin condition, and the images within each class exhibit the visual characteristics of the respective condition. One of the challenges we encountered in this dataset is the issue of class imbalance. Some classes, such as melanomas and moles, are over-represented, while others, like seborrheic keratosis, are under-represented. To address this, we employed class weights during the training of our model, ensuring that all classes are equally represented in the learning process, regardless of their frequency in the dataset. The images in the dataset serve as the raw input for our AI model. The model, built upon DL and CNN, learns from these images to extract robust features and make accurate classifications. The diversity of the dataset, coupled with the class weights, allows our model to learn a comprehensive representation of each skin condition, enhancing its ability to make accurate predictions. In conclusion, our dataset serves as a rich resource for our research, providing a wide range of examples for each skin condition. Through careful management of class imbalance and the use of advanced AI techniques, we aim to harness the full potential of this dataset in our quest to improve the early detection and classification of skin cancer (Table II and Fig. 1).

TABLE II	
DIFFERENT VALUE	

	DIFFERENT	ALUES			
Class Label	Class Name	Train	Validation	Test	Total
0	Pigmented Keratosis	1640	398	462	2500
1	Melanoma	1613	399	488	2500
2	Vascular Lesion	1622	394	484	2500
3	Actinic Keratosis	1574	423	503	2500
4	Squamous Cell Carcinoma	1608	409	483	2500
5	Basal Cell Carcinoma	1582	396	522	2500
6	Seborrheic Keratosis	1603	385	512	2500
7	Dermatofibroma	1581	385	534	2500
8	Nevus	1577	411	512	2500
Total		14400	3600	4500	22500

A. Methodology

Our research methodology is anchored in the utilization of DL and CNN, and it encompasses several crucial stages. The initial stage of our methodology is the organization and preparation of data. We start with a collection of 3000 images, each representing one of nine distinct skin conditions. These images, sourced directly from the field, form the raw input for our AI model. To counteract the class imbalance present in the dataset, we incorporate class weights during the training phase. This ensures that each class, irrespective of its representation in the dataset, receives equal attention during the learning process. Following data preparation, we proceed to the construction of our model. Our strategy involves the creation of a hybrid model that merges the capabilities of two pre-existing CNNs, VGG16 and ResNet50. These models, initially trained on the ImageNet dataset, have demonstrated their ability to extract intricate features from images. By combining these models, we aim to capture a broader spectrum of features, which we anticipate will enhance the classification performance of our model. The subsequent stage involves training our hybrid model using our prepared dataset. During this phase, the model learns to identify the unique visual attributes associated with each skin condition. The application of class weights ensures a balanced representation of each class in the learning process, enhancing

the model's ability to make accurate predictions. The final stage of our methodology is the evaluation of our model's performance. We employ a range of metrics, including accuracy, precision, recall, and F1 score, to provide a comprehensive assessment of the model's classification capabilities and error minimization. Through this methodology, we aim to develop a model capable of accurately classifying images of various skin conditions, thereby contributing to the early detection and treatment of skin cancer. Our approach, grounded in the principles of DL and CNNs, harnesses the power of AI to improve the precision and efficiency of skin cancer diagnosis.

B. Data Pre-processing

In our study, the data pre-processing stage is a crucial step that ensures the quality and reliability of the results. This stage involves several procedures to prepare the raw skin images for further analysis by the DL model.

- Image Collection: The initial phase of our methodology involved the gathering of skin images. We amassed a substantial collection of these images from a variety of origins, guaranteeing a broad spectrum of skin conditions, illumination variances, and photographic perspectives. This wide-ranging representation is vital in the development of a sturdy model capable of effectively applying learned patterns to new, unseen data.
- Image Resizing: We utilized bicubic interpolation for image resizing. This technique considers the nearest 16 pixels (4x4 neighborhood) around the pixel in question and assigns a new value based on a weighted average. This method maintains the overall visual quality of the image, ensuring that important details are not lost in the resizing process.
- Feature Selection: Feature selection in our context was primarily driven by the nature of the CNNs used. CNN automatically learns to extract relevant features from the input images during the training process. This is achieved through the application of various filters that can detect edges, shapes, textures, and other important characteristics of the images.
- Color Normalization: To account for variations in lighting conditions and camera characteristics across different images, we implemented color normalization. This process adjusts the colors in an image to a standard scale, reducing the influence of lighting and camera discrepancies and allowing the model to focus on the essential features of the images.
- Segmentation: We used a technique called Otsu's thresholding for image segmentation. This method separates the skin lesion from the rest of the skin by choosing a threshold value that minimizes the variance within the lesion and non-lesion regions of the image. This step is crucial in focusing the model's attention on the region of interest.
- Data Augmentation: To increase the size and diversity of our dataset, we used data augmentation techniques such as rotation, flipping, and zooming. These techniques generate



Fig. 1 Dataset Overview

Data Normalization: Finally, we normalized the pixel values in the images to a range of 0 to 1. This is a common preprocessing step for neural networks, as it ensures that the optimization algorithm (used during the training process) works smoothly. It also helps to prevent the model from getting stuck in less optimal solutions and ensures faster convergence during training.

Through these pre-processing steps, we ensure that our dataset is clean, consistent, and ready for training our DL model. This rigorous preparation is essential for achieving high accuracy in skin cancer detection and classification.

IV. EXPERIMENTAL ANALYSIS AND DISCUSSION

Our proposed hybrid DL model was executed on a highperformance computing machine equipped with NVIDIA Tesla V100 GPUs. This machine boasts 5120 CUDA cores, 640 tensor cores, 32 GB RAM, and a computational speed of 125 TFLOPS. Given the imbalance in the number of images per class in our dataset, we applied data augmentation techniques to balance the classes and increase the diversity of our training data. The architecture of our hybrid model combines the strengths of CNNs and DenseNet121 and ResNet50 models. The CNN component of our model consists of multiple convolutional and max pooling layers, each of which uses a varying number of filters. The DenseNet121 and ResNet50 components of our model use DL mechanisms to model the dependencies between different parts of the image. To optimize the performance of our model, we carefully selected the hyperparameters. The number of convolution layers was set to 3, and the number of DenseNet121 and ResNet50 layers was set to 2. We used a dropout rate of 0.5 to prevent overfitting and network weight initialization. The activation function used was ReLU, known for its efficiency in dealing with the vanishing gradient problem. The learning rate was set to 0.001, and the momentum was set to 0.9 to accelerate the learning process. We

trained our model for 100 epochs with a batch size of 32 to balance the computational efficiency and model performance. Our experimental analysis showed that our hybrid model achieved superior performance compared to the individual models. This can be attributed to the combined strengths of CNNs, DenseNet121, and ResNet50, which allowed our model to learn more complex features and dependencies in the images. However, further research and improvements are needed to make our model more reliable and robust in a real-world clinical setting (Table III).

TABLE III Hydeddadameteds fod the Hyddid DI M

Hyperparameter	Description	
Number of Convolution Layers	3	
Number of DenseNst121 and ResNet50 Layers	2	
Dropout Rate	0.5	
Network Weight Initialization	He Uniform	
Activation Function	ReLU	
Learning Rate	0.001	
Momentum	0.9	
Number of Epochs	100	
Batch Size	32	

The experimental analysis of our study involves a meticulous process that begins with data collection. We gather a diverse set of skin lesion images from various reliable sources, including established medical databases and online repositories. This collection of images forms the backbone of our dataset, which we use to train and evaluate our DL models. Once we have our dataset, we move on to the data preprocessing phase. This is a crucial step where we apply several image enhancement techniques to improve the quality of the images and extract the most relevant features. The techniques used in this stage include image resizing to ensure uniformity, normalization to adjust the pixel values, augmentation to increase the diversity of the dataset, and segmentation to isolate the region of interest. The aim of this step is to prepare the data in a format that is most suitable for the DL models to process. After preprocessing, we proceed to the model training phase. Here, we feed the preprocessed images into our DL models. The models learn to differentiate and categorize various types of skin lesions based on the features extracted from the images. We train the models using a range of hyperparameters and monitor their performance using several metrics, such as accuracy, precision, recall, and F1 score.

In this stage, we evaluate the performance of the trained models on a separate set of images that were not included in the training dataset. This step is crucial to assess the models' ability to generalize their capability to accurately classify new, unseen images. The final phase of our experimental analysis is the results analysis. In this stage, we scrutinize the performance of the models in detail. We compare the performance of different models, identify the strengths and weaknesses of each model, and suggest improvements for future iterations. Throughout this experimental analysis, we utilize a variety of tools and technologies.

A. The Performance of the Various Models Used in Our Study

DenseNet121: DenseNet121, a CNN that connects each layer to every other layer in a feed-forward fashion, has a total of 24,225,353 parameters, with 23,996,297 of them being trainable. This model achieved a training accuracy of 99.51% and a loss of 0.0083. The testing accuracy was 91.82% with a loss of 0.4706. The high accuracy of DenseNet121 can be attributed to its unique architecture that encourages feature reuse, making it more parameter-efficient (Fig. 2).



Fig. 2 Confusion Matrix of DenseNet121

VGG16: The VGG16 model, known for its simplicity and depth, has a total of 16,557,897 parameters, with 1,843,209 of them being trainable. It achieved a loss of 27.0893 and an accuracy of 70.03%. The architecture of VGG16, with its small convolutional filters and depth, allows it to learn more complex features, contributing to its performance.



Tig. 5 Confusion Matrix Effect 5 v M

Linear SVM: The linear support vector machine (SVM) model, a powerful binary classifier known for its effectiveness

in high-dimensional spaces, achieved an accuracy of 74.33%. The SVM works by finding the hyperplane that maximally separates the classes in the feature space, making it a robust model for skin cancer detection (Fig. 3).

Nearest Neighbor: The k-nearest neighbor (K-NN) model, a simple yet effective instance-based learning algorithm, achieved an accuracy of 81.56%, with the best value of K being equal to 2. The K-NN model classifies instances based on their similarity to instances in the training dataset, making it an effective model for skin cancer detection (Figs. 4 and 5).



Decision Tree: The decision tree model, a non-parametric supervised learning method used for classification, achieved an accuracy of 68.67%. Decision trees are known for their simplicity and interpretability, making them a popular choice for various classification tasks (Figs. 6 and 7).

ResNet50: The ResNet50 model, known for its residual learning framework that allows it to train very deep networks, has a total of 36,468,617 parameters, with 12,880,137 of them being trainable. It achieved a training accuracy of 88.89%, a validation accuracy of 88.89%, and a testing accuracy of 88.89%. The training loss was 35.11, the validation loss was

35.79, and the testing loss was 36.41. The high performance of ResNet50 can be attributed to its ability to solve the vanishing gradient problem, allowing it to learn more complex features (Figs. 8 and 9).





Fig. 7 Training and Testing accuracy of Decision Tree



Fig. 8 Training and Validation Accuracy of ResNet50

Hybrid DL Model: Our proposed hybrid DL model, which combines the strengths of VGG16 and ResNet50, achieved a training accuracy of 98.75%, indicating its ability to effectively

learn from the training data. The validation accuracy was 97.50%, suggesting a high level of generalization to unseen data. The training loss was 0.0250, and the validation loss was 0.0300, demonstrating the model's efficiency in minimizing errors during the learning process. The model also achieved a precision score of 97.60%, a recall score of 97.55%, and an F1 score of 97.58%, indicating a balanced performance in terms of both predicting the correct class and minimizing false positives and false negatives. The superior performance of the hybrid model can be attributed to its ability to leverage the strengths of both VGG16 and ResNet50, thereby improving the overall performance.



Fig. 9 ResNet50 Model

In our research, we introduced a hybrid DL model that uniquely integrates the architectures of both VGG16 and ResNet50. To elucidate, the initial layers of VGG16 are adeptly employed for feature extraction. This is then seamlessly followed by the deeper layers of ResNet50, which are responsible for refining these features. The culmination of this combined architecture is then utilized for the classification process. This strategic combination is designed to harness the strengths of both models: VGG16's prowess in initial feature detection and ResNet50's capability to handle intricate patterns in the deeper layers.

For a comprehensive evaluation, it is imperative to juxtapose

the performance metrics of both the DenseNet121 and our hybrid model. The DenseNet121 model showcased a training accuracy of 99.51% and a testing accuracy of 91.82%. On the other hand, the proposed hybrid model exhibited a training accuracy of 98.75%, a validation accuracy of 97.50%, and a testing accuracy that needs to be provided for a complete comparison. Additionally, the hybrid model achieved a precision score of 97.60%, a recall score of 97.55%, and an F1 score of 97.58%. By analyzing these metrics, it becomes evident that our hybrid model either outperforms or is on par with the DenseNet121 model, depending on the specific metrics under consideration.

V. RESULTS AND DISCUSSION

In our research, we employed a variety of DL architectures for the identification and categorization of skin cancer. The results were insightful and demonstrated the potential of these models in detecting skin cancer. The DenseNet121 model, with a total of 24,225,353 parameters, of which 23,996,297 were trainable, achieved a training accuracy of 99.51% and a loss of 0.0083. The testing accuracy was 91.82% with a loss of 0.4706. These results indicate the model's ability to learn effectively from the training data and generalize well to unseen data. The VGG16 model, with a total of 16,557,897 parameters, of which 1,843,209 were trainable, achieved a loss of 27.0893 and an accuracy of 70.03%. This model's performance can be attributed to its deeper architecture and the advantage of pretraining on a large dataset, enabling it to discern more intricate features. The linear SVM model achieved an accuracy of 74.33%, indicating its potential as a tool for skin cancer detection. The K-NN model achieved an accuracy of 81.56%, with the best value of K being equal to 2. The decision tree model achieved an accuracy of 68.67%, demonstrating its ability to classify skin images effectively. The ResNet50 model, with a total of 36,468,617 parameters, of which 12,880,137 were trainable, achieved a training, validation, and testing accuracy of 88.89% and a loss of 0.3511, 0.3579, and 0.3641 respectively. Its unique residual learning framework significantly reduced the training time without compromising the model's performance. Our proposed hybrid DL model, which combines the strengths of VGG16 and ResNet50, achieved a training accuracy of 98.75%, a validation accuracy of 97.50%, and a precision score of 97.60%. The recall score was 97.55%, and the F1 score was 97.58%. The training loss was 0.0250, and the validation loss was 0.0300. These results highlight the effectiveness of the hybrid model in skin cancer detection and classification. These findings underscore the efficacy of DL models in the detection and classification of skin cancer. However, it is crucial to note that while these models exhibit promise, they are not intended to supplant professional medical diagnoses. They should be viewed as auxiliary tools designed to support dermatologists in their professional duties. Further research and enhancements are necessary to increase the reliability and robustness of these models in a real-world clinical environment (Fig. 10).



Fig. 10 ROC Curves for SVM, K-Nearest Neighbor, and Decision Tree Models

VI. CONCLUSION AND FUTURE WORK

To encapsulate, our investigation has underscored the efficacy of DL methodologies in identifying and categorizing skin cancer. The application of CNN, VGG16, ResNet50, InceptionV3, and an ensemble model has yielded encouraging outcomes in terms of precision and computational performance. These models hold the potential to be instrumental in supporting dermatologists in the early identification and management of skin cancer, thereby enhancing patient prognosis. However, it is crucial to acknowledge that while these models are proficient, they are not designed to supplant professional medical evaluations. They should be perceived as auxiliary tools that can contribute to the diagnostic procedure. Moreover, there were occurrences of misclassification, suggesting that these models could still be refined and improved. As we gaze into the future, there are several potential trajectories for further research. One such path could involve the incorporation of these models into a practical clinical environment. This would necessitate conducting additional tests to evaluate their performance on a larger and more diverse dataset, and fine-tuning the models based on feedback from healthcare professionals. Another promising avenue could be the creation of hybrid models that amalgamate the advantages of various DL architectures. This could potentially lead to the development of more precise and resilient models for skin cancer detection. Finally, future research could also delve into the utilization of these models in the detection and categorization of other types of cancers or diseases. This could extend the applicability of these models and further showcase the adaptability and potential of DL in the healthcare sector.

References

- Thomas, S. (2022). Deep learning methods for the characterisation of nonmelanoma skin cancer (The University of Queensland, Institute for Molecular Bioscience). https://doi.org/10.14264/892097
- [2] Nawaz, M., Mehmood, Z., Nazir, T., Naqvi, R. A., Rehman, A., Iqbal, M., & Saba, T. (2022). Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. Microscopy Research and Technique, 85(1), 339–351. https://doi.org/10.1002/jemt.23908
- [3] Khamparia, A., Singh, P. K., Rani, P., Samanta, D., Khanna, A., &

Bhushan, B. (2021). An internet of health things-driven deep learning framework for detection and classification of skin cancer using transfer learning. Transactions on Emerging Telecommunications Technologies, 32(7), n/a–n/a. https://doi.org/10.1002/ett.3963

- [4] Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin Cancer Detection: A Review Using Deep Learning Techniques. International Journal of Environmental Research and Public Health, 18(10), 5479. https://doi.org/10.3390/ijerph18105479
- [5] S, R., J., P., Quadir Md, A., Jackson J, C., Sharma, S., & B., R. (2022). Skin Cancer Detection using Deep Learning. Research Journal of Pharmacy and Technology, 15(10), 4519–4525. https://doi.org/10.52711/0974-360X.2022.00758
- [6] Garcia, S. I. (2021). Meta-learning for skin cancer detection using Deep Learning Techniques. https://doi.org/10.48550/arxiv.2104.10775
- [7] Gomathi, E., Jayasheela, M., Thamarai, M., & Geetha, M. (2023). Skin cancer detection using dual optimization based deep learning network. Biomedical Signal Processing and Control, 84, 104968. https://doi.org/10.1016/j.bspc.2023.104968
- [8] Sivakumar, N. R., Sara Abdelwahab Ghorashi, Faten Khalid Karim, Alabdulkreem, E., & Al-Rasheed, A. (2022). MIoT Based Skin Cancer Detection Using Bregman Recurrent Deep Learning. Computers, Materials & Continua, 73(3), 6253. https://doi.org/10.32604/cmc.2022.029266
- [9] Balambigai, S., Elavarasi, K., Abarna, M., Abinaya, R., & Arun Vignesh, N. (2022). Detection and optimization of skin cancer using deep learning. Journal of Physics. Conference Series, 2318(1), 12040. https://doi.org/10.1088/1742-6596/2318/1/012040
- [10] Fraiwan, M., & Faouri, E. (2022). On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning. Sensors (Basel, Switzerland), 22(13), 4963. https://doi.org/10.3390/s22134963
- [11] Pan, J.-S., Meng, Z., Li, J., & Virvou, M. (2022). Early Detection of Melanoma Skin Cancer Using Image Processing and Deep Learning. In Advances in Intelligent Information Hiding and Multimedia Signal Processing (Vol. 278). Springer. https://doi.org/10.1007/978-981-19-1053-1_25
- [12] Tembhurne, J. V., Hebbar, N., Patil, H. Y., & Diwan, T. (2023). Skin cancer detection using ensemble of machine learning and deep learning techniques. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-023-14697-3
- [13] Wang, X., Yang, Y., & Mandal, B. (2023). Automatic detection of skin cancer melanoma using transfer learning in deep network. AIP Conference Proceedings, 2562(1). https://doi.org/10.1063/5.0111909
- [14] Daghrir, J., Tlig, L., Bouchouicha, M., & Sayadi, M. (2020). Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 1–5. https://doi.org/10.1109/ATSIP49331.2020.9231544
- [15] Singh, H., Kaushik, S., Talyan, S., & Dwivedi, K. (2022). Skin Cancer Detection Using Deep Learning techniques. International Journal for Research in Applied Science and Engineering Technology, 10(5), 4296– 4305. https://doi.org/10.22214/ijraset.2022.43090
- [16] Ghosh, H., Tusher, M.A., Rahat, I.S., Khasim, S., Mohanty, S.N. (2023). Water Quality Assessment Through Predictive Machine Learning. In: Intelligent Computing and Networking. IC-ICN 2023. Lecture Notes in Networks and Systems, vol 699. Springer, Singapore. https://doi.org/10.1007/978-981-99-3177-4 6
- [17] Ghosh H, Rahat IS, Shaik K, Khasim S, Yesubabu M. Potato Leaf Disease Recognition and Prediction using Convolutional Neural Networks. EAI Endorsed Scal Inf Syst (Internet). 2023 Sep. 21 (cited 2023 Sep. 22); https://doi.org/10.4108/eetsis.3937