International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Intelligent Systems for Skin Cancer Diagnosis Using Machine Learning

A. Jothi[1], D. Manju[2], R. Geethamani[3]

Assistant Professor[1&3], II MSC[2]
Department of Computer Science
K. R. College of Arts & Science, K. R. Nagar, Kovilpatti- Tamil Nadu 628503

ABSTRACT

Skin cancer is one of the most common types of cancer worldwide, with increasing rates of incidence that present significant challenges for healthcare systems. Early detection and precise diagnosis are vital for effective treatment and improving patient outcomes, underscoring the need for innovative diagnostic methods. Recently, machine learning (ML) algorithms have proven to be powerful tools in analyzing medical imaging data, providing critical assistance to healthcare professionals in diagnosing skin cancer. This review paper provides an in-depth exploration of ML classification algorithms specifically designed for skin cancer detection and diagnosis. It covers various forms of skin cancer, such as melanoma, basal cell carcinoma, and squamous cell carcinoma, emphasizing their unique features and the diagnostic hurdles they present. The paper also discusses current advanced ML techniques, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, and Convolutional Neural Networks (CNN), evaluating their strengths and limitations in the context of skin cancer classification. A thorough search was conducted across several academic databases, and the dataset utilized is the International Skin Imaging Collaboration (ISIC). The continuous advancements in skin cancer classification through machine learning offer promising prospects for improving diagnostic precision and facilitating more personalized treatment options, ultimately enhancing patient care and outcomes.

Keyword: KNN, SVM, CNN, RANDOM FOREST

I. INTRODUCTION

Skin cancer is one of the most prevalent cancers globally, with rising incidence rates posing significant challenges to healthcare systems. Early detection and accurate diagnosis are crucial for effective treatment and improving patient outcomes, making the exploration of innovative diagnostic methods essential. Recently, machine learning (ML) algorithms have emerged as transformative tools in the analysis of medical imaging data, providing valuable support to clinicians in diagnosing skin cancer. This review paper aims to offer a comprehensive overview of ML classification algorithms tailored for skin cancer detection and diagnosis. It delves into various types of skin cancer, including melanoma, basal cell carcinoma, and squamous cell carcinoma, highlighting their distinct characteristics and the diagnostic challenges they present. Furthermore, the paper examines current state-of-the-art ML techniques, such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest and Convolutional Neural Networks (CNN), discussing their strengths and limitations in the context of skin cancer classification. A systematic search was conducted across multiple academic databases. The dataset we used is International Skin Imaging Collaboration (ISIC).The ongoing advancements in skin cancer classification using machine learning hold the promise of enhancing diagnostic accuracy and enabling more personalized treatment approaches, ultimately contributing to better patient care and outcomes**.** The ongoing advancements in skin cancer classification using machine learning hold the promise of enhancing diagnostic accuracy and enabling more personalized treatment approaches, ultimately contributing to better patient care and outcomes. By comparing KNN, CNN, SVM and Random Forest models CNN provide the best accuracy.

II. LITERATURE SURVEY

The integration of artificial intelligence (AI) in dermatology, particularly for skin cancer diagnosis, has shown promise in improving early detection and reducing the workload on specialists. Machine learning (ML), a subset of AI, has been instrumental in this progress, utilizing algorithms to analyze dermoscopic and clinical images. With deep learning techniques, especially convolutional neural networks (CNNs), systems can accurately classify skin lesions into categories such as benign and malignant. Despite these advancements, challenges remain, including the need for large, diverse datasets, and the protection of patient privacy. Continued collaboration and research are essential for refining AI tools and enhancing their clinical application, while ensuring they complement, rather than replace, dermatological expertise.[1] The literature highlights the increasing global incidence of skin cancer, particularly melanoma, and emphasizes the importance of early detection and accurate classification of various skin cancer types. Techniques such as image enhancement, segmentation, and advanced classifiers like MSVM demonstrate promising results in improving diagnostic accuracy, achieving around 96.25% in identifying different skin cancer types [2].skin cancer diagnosis due to their effectiveness in handling high-dimensional data and creating optimal decision boundaries. Research indicates that SVM can accurately differentiate between malignant melanoma and benign lesions, often achieving high classification accuracy rates exceeding 90%. Studies emphasize the importance of feature selection and preprocessing steps, as well as the kernel trick, which enables SVM to manage non-linear relationships in data. Challenges include the need for large, well-annotated datasets to train robust models and the interpretability of results in clinical settings. Ongoing research aims to enhance SVM applications by integrating multi-modal data and improving model transparency. Overall, SVM remains a valuable tool in the evolving landscape of skin cancer diagnosis[3].In Saba (2020), the author examines the advancements in cancer detection through machine learning, focusing on algorithms such as Convolutional Neural Networks (CNNs). The paper highlights how CNNs have significantly improved the accuracy of tumor detection in medical imaging, particularly in identifying malignant lesions in skin, lung, and breast cancers. Saba emphasizes the strength of CNNs in automatically extracting features from images, which enhances diagnostic efficiency and reduces reliance on manual inter pretation. The author also addresses the challenge so implementing CNNs, including the need for large, well-annotated datasets and the potential for overfitting. Additionally, the review discusses the importance of model validation across diverse populations to ensure generalizability. Overall, Saba's work underscores the transformative potential of CNNs in the early detection and diagnosis of various cancers [4].The largest organ of the human body is the skin, which serves as a protective barrier, regulates temperature, and enables sensory functions. Skin lesions, which can develop due to infections or mishandling of pre-existing lesions, are a major concern, as they may lead to skin cancer, including types like basal cell carcinoma, squamous cell carcinoma, and melanoma. Early detection is crucial for improving survival rates, and automatic detection methods, including segmentation and feature extraction using techniques like ABCD rule, GLCM, and HOG, offer more reliable and accurate results than manual inspection[5].This literature review underscores the rising incidence of skin cancer, particularly melanoma and non- melanoma types, emphasizing the critical importance of early detection for effective treatment. It discusses the limitations of traditional visual inspections by dermatologists compared to the enhanced diagnostic capabilities offered by dermoscopy, which improves diagnostic accuracy by 10-30%. The review highlights automated image processing techniques, detailing a systematic approach involving lesion segmentation, feature extraction, and classification. K-means clustering is identified as a key segmentation method, while Local Binary Patterns (LBP) combined with color percentiles are used for feature extraction. The performance of various classifiers, particularly Support Vector Classifier and K- NN, is evaluated, revealing their effectiveness in distinguishing skin cancer images. Overall, the paper illustrates the potential of automated systems to significantly improve skin cancer detection and outcomes [6].Skin cancer remains a significant global health concern, with a rising number of cases each year, necessitating early detection to improve survival rates. Traditional diagnostic methods involve visual inspection, dermoscopy, and biopsy, which can be time-consuming and are highly dependent on the clinician's skill. Recent advancements in computer vision and deep learning haveled to the development of auto mated skin cancer detection tools. Studies have demon strated the effectiveness of Convolutional Neural Networks (CNNs) for classifying skin lesions, significantly improving diagnostic accuracy. Models like Google Net Inception and Alex Net have set benchmarks by achieving high accuracy rates. Furthermore, newer approaches leverage non-parametric methods, avoiding the limitations of traditional algorithms that require normally distributed data, thereby enhancing the reliability of automated diagnosis [7].Skin cancer, particularly melanoma, is a leading cause of cancer-related deaths, making accurate diagnosis essential. This paper proposes a novel method combining machine learning and deep learning techniques for skin cancer detection, achieving 93% accuracy and high recall scores, outperforming expert dermatologists and existing methods, and offering a valuable tool to prevent misdiagnosis [8].The literature on skin cancer detection emphasizes the increasing prevalence of skin cancers, particularly melanoma, due to environmental factors like UV radiation. Recent studies focus on advanced image processing techniques for analyzing dermoscopy images, utilizing classifiers such as Support Vector Machines (SVM), and Naive Bayes. These methodologies aim to improve diagnostic accuracy, sensitivity, and specificity in distinguishing between malignant and benign lesions. Research highlights the significant role of genetic predispositions and lifestyle choices, such as tanning bed use, in skin cancer risk. Furthermore, there is a consensus on the importance of early detection through visual examinations and innovative computational tools in enhancing treatment outcomes. Overall, this body of work underscores the need for integrated approaches combining clinical expertise and technology in combating skin cancer [9].The study explores the benefits of combining human expertise and artificial intelligence (AI) for skin cancer classification. Previous research often positioned AI, specifically convolutional neural networks (CNNs), against dermatologists, showing that CNNs could outperform humans in distinguishing melanoma from benign lesions. However, the concept of combining classifiers, commonly used in machine learning, was applied here to see if merging the decisions of CNNs and dermatologists could improve diagnostic accuracy. By utilizing a dataset of over 11,000 dermoscopic images, including from the widely recognized HAM10000 Dataset, this study demonstrated that a hybrid approach (combining CNN and dermatologists) outperformed each classifier on its own. The findings highlight the potential for improved diagnostic outcomes in clinical settings when human intelligence is augmented with AI [10].

III. METHODOLOGY

In this classification model KNN, CNN, SVM and Random Forest algorithms are used. After storing the training dataset into a selected directory with class labels, a path call function is used to save the dataset into the model. In this proposed model, four classification algorithms are used to classify skin disease. The first step of the proposed model is to include training image data into the system. Then a pre-processing work is done into the system to prepare the image dataset and train the data to develop the classification system. After that, testing data of every class are entered into the system, as shown in Fig.1.

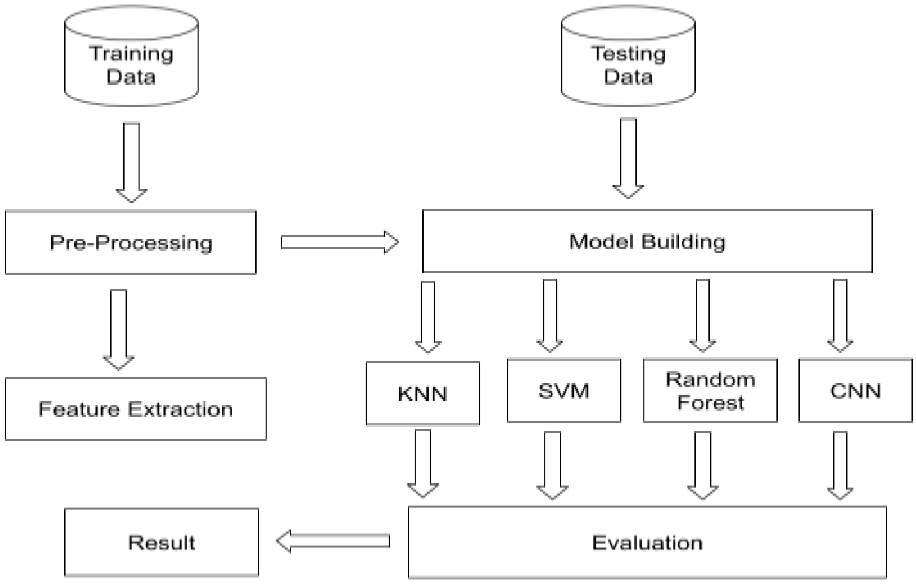

Fig 1. Architecture of proposed model

The testing image of every class goes into KNN , CNN, SVM and Random Forest classifier to classify the disease class. After feature extraction and evaluation system shows the expected class label name along with accuracy.

3.1 DATASET

The International Skin Imaging Collaboration (ISIC) dataset is taken from Kaggle, which contains various skin cancer images differentiated by three categories: “Melanoma”, ” Basal cell carcinoma” and “squamous cell carcinoma”. The training dataset for this proposed model is made up of 7928 skin disease pictures with 3 class labels that were obtained from an open-source benchmark dataset. For testing purposes,100 images for each class label are stored in the different directory to test the model.

Fig 2. Starting Stage of Skin Cancer lesion images

(a) Basal Cell (b)Squamous Cell (c)Melanoma

Fig 3. Types of Skin Cancer Lesion images

3.2. PROPOSED METHODOLOGY

In this proposed model CNN, SVM, KNN and Random Forest algorithms are used to classify skin disease.

a) K-Nearest Neighbors

KNN is a widely used non-parametric machine learning algorithm, and it is a supervised technique for data learning . It is used for both regression and classification. KNN algorithm is mainly used for handwritten detection, image classification, and recognition. In KNN, k represents the nearest neighbor point which is responsible or closely matched with the testing data class.

b) Random Forest Algorithm

The RF algorithm is one of the broadly used supervised machine learning algorithms. It makes decision trees for building the forest that can classify. Bagging is the training method of the RF algorithm. Bagging is the combination of total learning models that impacts the overall performance of the system .The main advantage of random forest that it can be used in regression and classification problems. RF is nearly the same as bagging or decision tree classifier. RF algorithm consists of multiple decision trees with N features. In regression problems, the mean squared error (MSE) rate is an important parameter in the RF.

c) Support Vector Machine (SVM)

SVM (Support Vector Machine) is a supervised machine learning algorithm which is mainly used to classify data into different classes. Unlike most algorithms, SVM makes use of a hyperplane which acts like a decision boundary between the various classes. SVM can be used to generate multiple separating hyperplanes such that the data is divided into segments and each segment contains only one kind of data.

d) Convolution Neural Network

CNNs are neural networks with a specific architecture that have been shown to be very powerful in areas such as image recognition and classification. CNNs have been demonstrated to identify faces, objects, and traffic signs better than humans and therefore can be found in robots and self-driving cars. CNNs are a supervised learning method and are therefore trained using data labeled with the respective classes. Essentially, CNNs learn the relationship between the input objects and the class labels and comprise two components: the hidden layers in which the features are extracted and, at the end of the processing, the fully connected layers that are used for the actual classification task.

Pseudo-code

Step 1: Preprocess the Dataset

- Apply image augmentation to increase dataset diversity, resize images to a fixed size (e.g., 224x224), and normalize pixel values using preprocessing functions like preprocess_input to prepare the data for the model.

Step 2: Build CNN Model Using Pretrained EfficientNetB0

- Load the EfficientNetB0 model with pre-trained ImageNet weights, remove its top classification layers, and flatten the output to prepare it for the custom classifier.

Step 3: Add Custom Layers

- Add a fully connected layer (Dense) with 128 neurons and ReLU activation, followed by a Dropout layer to prevent overfitting, and a final Dense layer with a softmax activation for multi-class classification.

Step 4: Compile the Model

- Compile the model by selecting the Adam optimizer, categorical cross-entropy loss for multi-class classification, and accuracy as the evaluation metric to track model performance.

Step 5: Train the Model

- Train the model using the training data generator, specify the number of epochs and batch size, monitor validation accuracy and loss during training, and evaluate the model after training on the test/validation set.

IV. EXPERIMENTAL RESULTS

The proposed method is applied to skin lesion images collected from ISIC. The datasets consist of 7928 skin disease images.

The process begins with preprocessing the dataset, which involves applying image augmentation to increase the diversity of the data, resizing images to a fixed size (e.g., 224x224 pixels), and normalizing pixel values using functions like preprocess_input to prepare the images for the model. This ensures that the model receives high-quality, consistent data. In the next step, a CNN model is built using the pre-trained EfficientNetB0 model, which has been trained on the ImageNet dataset. The top classification layers of EfficientNetB0 are removed, and the output is flattened to make it compatible with the custom classifier.

The model is then enhanced with custom layers, including a fully connected dense layer with 128 neurons and ReLU activation, a Dropout layer to prevent overfitting, and a final softmax layer for multi-class classification. The model is compiled with the Adam optimizer, categorical cross-entropy loss (for multi-class classification), and accuracy as the evaluation metric. In the final step, the model is trained on the dataset using a data generator, specifying the number of epochs and batch size. The model's performance is monitored through validation accuracy and loss, and after training, the model is evaluated on a separate test or validation set to assess its overall performance.

Table 1 Accuracy of models

S.NO	Algorithm	Accuracy
1	SVM	65%

2	KNN	75%
3	RANDOM FOREST	80%
4	CNN (EfficientNetB0**)**	90%

In the comparison of four machine learning algorithms—Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, and Convolutional Neural Networks (CNN)—for skin cancer detection, the CNN achieved the highest accuracy of 90%. CNNs excel in processing complex image data, making them the best choice for tasks like skin cancer detection, where image classification is crucial. Random Forest, with an accuracy of 80%, performed better than both KNN (75%) and SVM (65%), although it still lags behind CNNs in handling image-based tasks. While Random Forest is an effective ensemble method, particularly for tabular data, it does not match the performance of CNNs when it comes to extracting features from images. KNN, a simple and intuitive algorithm, performed decently but struggled with larger and more complex datasets, and SVM, despite its strong generalization abilities, was the least effective for this particular task. Overall, the results highlight the superiority of CNNs for skin cancer detection, demonstrating how advanced techniques can significantly improve diagnostic accuracy compared to traditional machine learning algorithms

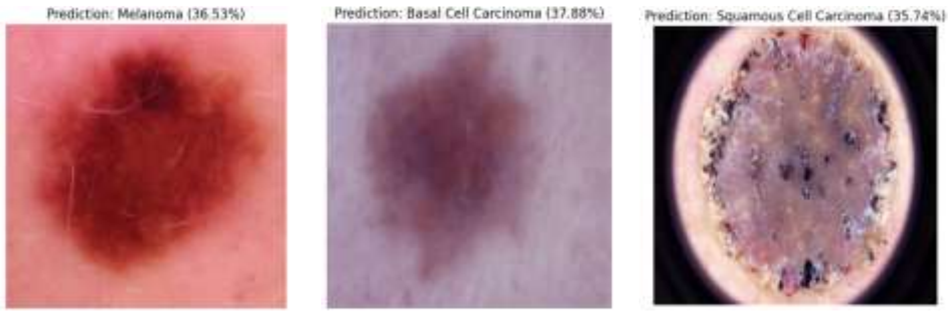

Fig 4: Types Skin Cancer Prediction

Table 2 ***Multi classification of skin cancer***

S.No	**Classification**	**Predicted Accuracy**
1	Melanoma	36.53%
2	BasalCell Carcinoma	37.88%
3	SquamousCell Carcinoma	35.74%

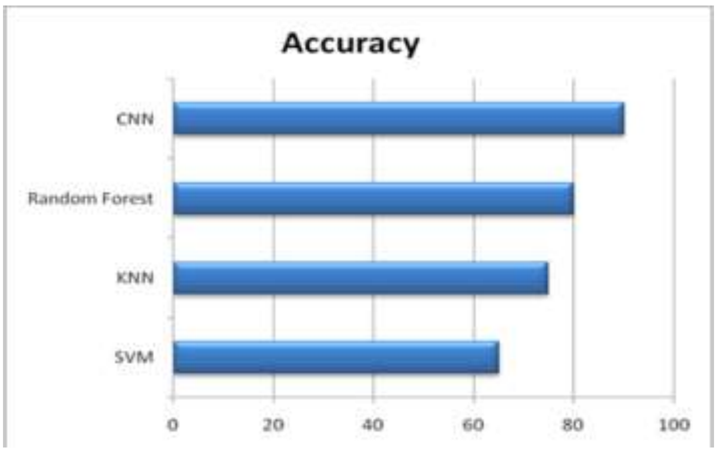

Fig 5. Accuracy of models

To build a skin cancer detection model using EfficientNetB0, the first step involves preprocessing the dataset. This includes applying image augmentation to increase data diversity and improve model generalization, resizing images to a fixed size of 224x224 pixels (the expected input size for EfficientNetB0), and normalizing pixel values using preprocess_input to scale them between 0 and 1. Next, the pre-trained EfficientNetB0 model is loaded with weights from ImageNet, and its top classification layers are removed, making the model ready for custom modifications. A GlobalAveragePooling2D layer is applied to reduce the output dimensions, followed by a Dense layer with 128 neurons and ReLU activation for feature learning. A Dropout layer is added to help prevent overfitting, and a final Dense layer with softmax activation is used for multi-class classification based on the number of classes in the dataset.

The model is then compiled using the Adam optimizer, categorical cross-entropy loss, and accuracy as the evaluation metric. During training, the model uses a data generator for both the training and validation datasets, with image augmentation applied only to the training data. The model is trained for a predefined number of epochs, monitoring both training and validation accuracy. Optionally, the model can be fine-tuned by unfreezing the layers of EfficientNetB0 after initial training, using a lower learning rate to fine-tune the pre-trained layers for better performance. This approach leverages the power of EfficientNetB0's pre-trained features, making it a robust solution for skin cancer detection.

V. CONCLUSION

This study compares the performance of four different machine learning models—SVM, KNN, CNN, and Random Forest—on the same dataset to develop a more effective method for detecting skin cancer. Using the International Skin Imaging Collaboration (ISIC) dataset for experimental analysis, Convolutional Neural Networks (CNNs) achieved the highest accuracy of 90% for skin cancer detection, outperforming traditional machine learning algorithms such as Random Forest (80%), K-Nearest Neighbors (75%), and Support Vector Machines (65%). CNNs excel at processing complex image data, making them a powerful tool for diagnostic applications. While each algorithm has its own strengths and weaknesses, the integration of advanced techniques like CNNs marks a significant step forward in creating more precise and efficient diagnostic tools. These innovations not only enhance diagnostic accuracy but also support personalized treatment strategies, ultimately improving patient care and outcomes in the battle against skin cancer.

VI. REFERENCES

1. Das, K., Cockerell, C. J., Patil, A., Pietkiewicz, P., Giulini, M., Grabbe, S., & Goldust, M. (2021). Machine learning and its application in skin cancer. International Journal of Environmental Research and Public Health, 18(24), 13409. [Google Scholar]

2. Monika, M. K., Vignesh, N. A., Kumari, C. U., Kumar, M. N. V. S. S., & Lydia, E. L. (2020). Skin cancer detection and classification using machine learning. Materials Today: Proceedings, 33, 4266-4270. [Google Scholar]

3. Murugan, A.,Nair,S. A.H., Preethi, A.A. P., &Kumar,K.S.(2021).Diagnosisof skincancer using machine learning techniques. Microprocessors and Microsystems, 81, 103727.[Google Scholar]

4. Saba, T. (2020). Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. Journal of infection and public health, 13(9), 1274-1289.[Google Scholar]

5. Vidya, M., & Karki, M. V. (2020, July). Skin cancer detection using machine learning techniques. In 2020 IEEE international conference on electronics, computing and communication technologies (CONECCT) (pp. 1-5). IEEE..[Google Scholar]

6. Anas, M., Gupta, K., & Ahmad, S. (2017). Skin cancer classification using K-meansclustering. International Journal of Technical Research and Applications, 5(1), 62-65.[Google Scholar]

7. Kadampur, M. A., & Al Riyaee, S. (2020). Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermalcell images. Informatics in Medicine Unlocked, 18, 100282.[Google Scholar]

8. 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, 82(18), 27501-27524.[Google Scholar]

9. Janney, B. J., Roslin, S. E., & Shelcy, M. J. (2018, July). A comparative analysis of skin cancer detection based on naive bayes classifier. In2018 International Conference onRecent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE) (pp. 1694-1699). IEEE.[Google Scholar]

10.Hekler,A.,Utikal,J.S.,Enk,A.H.,Hauschild,A.,Weichenthal, M.,Maron,R. C. &Thiem,A. (2019). Superior skin cancer classificationby the combination of human and artificial intelligence. European Journal of Cancer, 120,114-121.[Google Scholar]