

Deep Learning for Skin Cancer Classification: A Comparative Study of CNN and Vgg16 on HAM10000 Dataset

Yashwant S. Ingle¹, Dr. Nuzhat Faiz Shaikh²

¹Research Scholar, Dept. of Computer Engineering, SKNCOE Research Center, Pune, India

²PhD Research Supervisor, Dept. of Computer Engineering, SKNCOE Research Center, Pune, India

Article History:

Received: 18-04-2024

Revised: 09-06-2024

Accepted: 27-06-2024

Abstract:

Skin cancer is one of the most dangerous types of cancer among the cancers. The early detection of skin cancer helps resolve it. Hence, it is necessary to diagnose the disease as early as possible. This paper presents Convolutional Neural Networks and the Vgg16 algorithm to recognize skin cancer types. The HAM10000 dataset, which comprises seven distinct forms of skin cancer, melanocytic nevi (nv), Melanoma (mel), basal cell carcinoma (bcc), actinic keratoses (akiec), vascular lesions (vasc), and dermatofibroma (df). This system aims to improve classification accuracy; the methodology necessitates extensive dataset preparation, including scaling, normalization, and augmentation. The Vgg16 algorithm, when combined with the CNN architecture, offers a robust basis for the classification of skin cancer. Comprehensive details on regularization techniques, optimization strategies, and training parameters are included to ensure openness and reproducibility. The system's performance is evaluated using accuracy, precision, recall, and F1-score for every type of skin cancer. This paper highlights the usefulness of the proposed method in skin cancer diagnosis and looks at challenges, constraints, and prospects for further research. New methods for identifying skin cancer are being developed with the help of this research, which can improve patient outcomes and clinical decision-making.

Keywords: Skin cancer recognition, CNN, Vgg16, HAM10000 dataset, Image classification.

1. INTRODUCTION

With a rising incidence and substantial morbidity, skin cancer is a serious worldwide health problem. The World Health Organisation reports that 1 in 3 tumors detected globally are skin malignancies, highlighting the critical need for cutting-edge diagnostic methods to facilitate early diagnosis and treatment. The three most common skin cancer types are Melanoma, basal, and squamous cell carcinoma. The conditions present a significant challenge to healthcare systems and emphasize the necessity for precise and effective categorization techniques.

Early detection of skin cancer is important for achieving treatment on time. Early detection methods successfully treat early-stage skin cancer, reducing the impact on patients' lives and minimizing healthcare expenses. If research is delayed, cancer can spread to more severe stages. Hence, more intense therapies are needed, perhaps lowering survival chances. Creating accurate and effective diagnostic methods is necessary to enhance early detection rates and improve patient prognosis.

This study aims to implement the Vgg16 algorithm and CNN to create a reliable system for diagnosing skin cancer. This approach needs to improve the accuracy and efficiency of skin cancer

classification by utilizing the vast information on different types of skin cancer included in the HAM10000 dataset. The proposed method improves current knowledge by addressing specific skin lesion classification challenges.

The study evaluates the suitability of the proposed strategy in real clinical circumstances. Understanding the advantages and disadvantages of the current model is crucial for incorporating it into healthcare operations. The research aims to enhance patient experiences and diagnostic precision by integrating advanced deep-learning techniques with practical applications in dermatology.

2. LITERATURE SURVEY

Ammara Masood et al. [1] presented data and conclusions from the most important documented implementations. They analyzed the outcomes and assessed the effectiveness of many classifiers to diagnose skin lesions. Details on several variables that affect the approach's performance are given where feasible. They evaluate the results based on these models and offer a framework for comparing the diagnostic models for skin cancer. There is an emphasis on the limits of certain recent studies, and suggestions for further study are made.

Nazia Hameed et al. [2] reviewed the latest advancements in computer-aided diagnosis systems and analyzed new approaches in various stages of these systems. Analyses and reports are conducted on statistics and outcomes of significant and recent implementations. Based on several factors, including accuracy, dataset, computing time, color space, and machine learning approach, they assessed the performance of the most current study. The data were organized into tables to support burgeoning researchers in computer-aided skin diagnostic systems. Additionally, research issues about different parts of computer-aided skin cancer diagnostic systems are highlighted.

A CAD approach to the detection and categorization of skin lesions was presented by Maciej Ogorzałek et al. [3]. It integrates medical knowledge with modern technologies, including statistical learning, pattern classification, image processing, and ensembling techniques. In 98% of the cases, their method has produced accurate classification.

In their review of the current state of automated skin lesion identification, Fabio Santos et al. [4] thoroughly examined the challenges and opportunities facing dermatological care.

Shetu Rani Guha et al. [5] described a convolutional neural network (CNN), a machine-learning technique for categorizing seven skin conditions. The 2018 International Skin Imaging Collaboration (ISIC) dataset's classification accuracy was increased using CNN and transfer learning. Studies show that transfer learning improves accuracy by 11% compared to CNN alone. Compared to other approaches now in use, the performance of this proposed strategy shows promise.

Mengistu and colleagues (2016) introduced a digital image processing method for identifying and forecasting various forms of skin cancer through digital image processing techniques. The established classifications of skin cancer kinds managed the categorization system. ANN, KNN, and Naïve Bayes classifiers are not as effective in detecting and classifying skin cancer as integrating Self-Organizing Map (SOM) and Radial Basis Function (RBF). Also, it was shown that morphology and color features had stronger discrimination power than texture features; nevertheless, classification accuracy increased when morphology, texture, and color features were combined.

Uzma Bano Ansari and colleagues introduced a skin cancer detection method utilizing a support vector machine (SVM) for early skin cancer identification. The diagnostic method uses the Support Vector Machine (SVM) algorithm and image processing methods. To improve the image and eliminate noise, the dermoscopy image of skin cancer goes through several pre-processing steps. The image is segmented using the Thresholding method. Certain characteristics of the image need to be isolated using the GLCM approach. The classifier receives these features as input. SVM is utilized for classification. It categorizes the provided image as either malignant or non-cancerous.

Enakshi Jana and colleagues conducted a thorough review of existing technologies designed for skin cancer detection and performed a precise comparison of cutting-edge algorithms in this field. To diagnose skin cancer, every available technology is thoroughly reviewed. AdaBoost and SVM are the best methods for identifying skin cancer. An overview and analysis of the various ANN architectural configurations and the performance and accuracy outcomes are provided. SVM is also used to categorize skin cancer images. A concise explanation of how Melanoma works and how it can be detected is provided, aiding in the differentiation between normal and malignant skin cells.

A semi-supervised learning system for automatic melanoma detection using dermoscopic images was presented by Ammara Masood et al. [9]. Labeled and unlabeled data are used to construct a deep belief architecture, then refined with an exponential loss function to improve the labeled data's differentiation. Simultaneously, a self-advised SVM method enhances classification accuracy by reducing the influence of incorrectly classified data. To improve redundancy and generalization ability, bootstrapped Deep network models and SA-SVMs based on polynomial and radial basis functions are trained utilizing random training samples. The least squares estimate weighting is used to mix the data. One hundred dermoscopic pictures are used to assess the model. The performance of the suggested model utilizing deep brain processing surpasses popular classification approaches such as KNN, ANN, SVM, Expectation maximization, and transductive SVM.

Vijayalakshmi M M et al. [10] introduced an entirely automated technique for recognizing dermatological diseases using lesion photos, which is a machine-based approach rather than relying on traditional medical professionals for detection. The three stages of the model's construction are development, prediction, and data collection and augmentation. With many AI methods, including convolutional neural networks, support vector machines, and image processing software, we developed an enhanced framework that produced an accuracy of 85%.

3. DEEP LEARNING ALGORITHM

The system utilizes the CNN and Vgg16 algorithms for skin cancer identification and classification. A detailed explanation of the CNN and Vgg16 algorithms is provided in this section.

A. CNN

Fig. 1 shows the architecture of the CNN algorithm used to classify skin cancer.

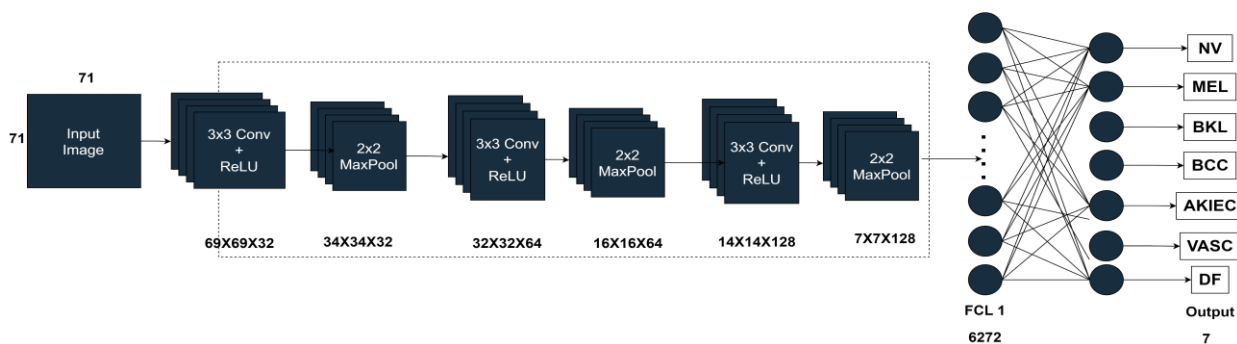


Fig. 1. Architecture of CNN

Fig. 1 displays a specialized CNN model created to identify skin cancer. The Keras library's Sequential API is utilized for model generation. Three color channels (RGB) and 71 pixels in width and height are required for the input images. The first layer consists of a Rectified Linear Unit (ReLU) activation function and a Conv2D layer with 32 filters, each measuring (3, 3). This layer is the primary feature extractor, gathering basic characteristics from the incoming photos. The convolutional layer is followed by a MaxPooling2D layer with a pool size of (2, 2) to reduce the spatial dimensions of the feature maps while keeping important information. Three more sets of Conv2D and MaxPooling2D layers are then incorporated into the model, enhancing the sophistication of acquired features. The second Conv2D layer has 64 filters, the third layer has 128 filters, and the final Conv2D layer has 128 filters. Every convolutional layer is paired with a MaxPooling2D layer to aid hierarchical feature extraction and downsampling.

A flattened layer is applied after the convolutional layers to transform the multi-dimensional feature maps into a one-dimensional vector to prepare the data for the dense layers. Two dense layers are given the compressed representation. The initial dense layer comprises 512 neurons and uses the ReLU activation function to introduce non-linearity and learn more advanced characteristics. A Dropout layer with a dropout rate of 0.5 is inserted following the initial dense layer to prevent overfitting. The last dense layer consists of seven neurons, each representing one of the dataset's seven distinct skin cancer classes. To ensure the output represents a probability distribution across the classes, the layer uses the softmax activation function. The model is intended for multi-class classification, with each class corresponding to one of the seven kinds of skin cancer.

B. Vgg16

Fig. 2 displays the architectural design of the Vgg16 algorithm used for skin cancer classification.

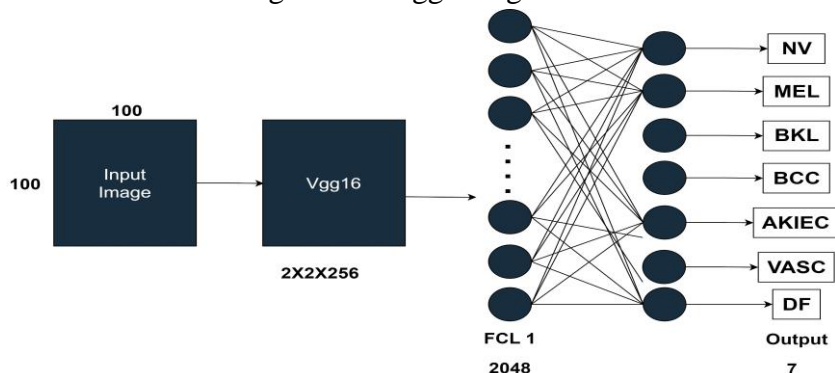


Fig. 2. Architecture of Vgg16

Fig. 2 illustrates a skin cancer detection method employing the Vgg16 model. The following code outlines the structure of the skin cancer recognition model. A Sequential model is generated to enable a linear arrangement of layers. The Vgg16 basic model is the first layer to function as a feature extractor. After the convolutional layers, a Flatten layer is added to convert the output into a one-dimensional array. Flattening is crucial for the following completely connected layers. Next, a dense layer consisting of 64 neurons and utilizing a rectified linear unit (ReLU) activation function is included to introduce non-linearity to the model. A Dropout layer with a dropout rate 0.2 reduces overfitting by randomly deactivating 20% of the neurons during training. The output layer is a dense layer with seven neurons and a softmax activation function that produces probability distributions for each of the seven forms of skin cancer.

4. PROPOSED SYSTEM

The block diagram of the proposed skin cancer recognition system is presented in Fig.3.

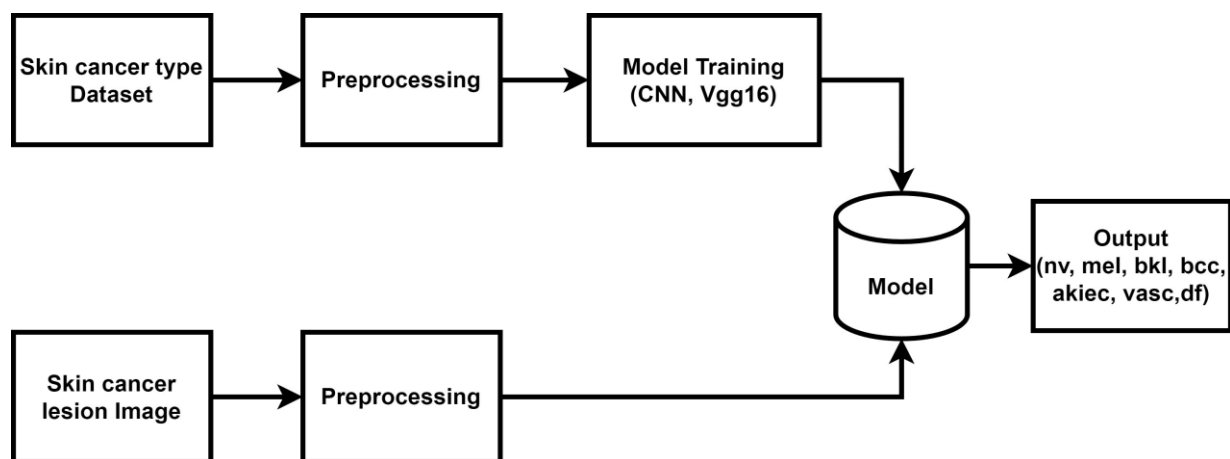


Fig. 3. Block diagram of the proposed system

A. Dataset

This system makes use of the HAM1000 dataset [11]. Carefully selected dermatoscopy photographs from a range of demographics were obtained using a variety of techniques for collection and preservation. 10,015 dermatoscopic images comprise the carefully selected dataset, intended to serve as a comprehensive training set for academic machine-learning research. Important diagnostic characteristics associated with pigmented lesions are included in every image in the collection. The dataset was carefully selected for the machine-learning algorithms utilized in dermatological diagnostics.

B. Pre-processing

The pre-processing stage of the skin lesion classification pipeline involves importing dermatoscopic images and their subsequent resizing to 71x71 pixels. This is done in order to ensure that the dataset is consistent throughout. Several data augmentation techniques increase the dataset's variety, including random flips and rotations. Normalized pixel values are also utilized. Converting the color space to RGB and removing elements that are not helpful, such as background components, are both necessary steps in the process of improving the input data.

C. Model Training

Skin lesion classification models are trained using Vgg16 architecture and CNN. Identifying distinctive features in the incoming data means methodically modifying the model's parameters. The initial framework is pre-trained, utilizing convolutional layers intended for visual information extraction based on the Vgg16 model. The model can become more adept at identifying skin lesions through transfer learning by utilizing input from several images in the ImageNet dataset. By employing additional thick layers created on top of the Vgg16 basis, the model can distinguish between different skin lesions, improving task-specific learning. During training, batches of pre-processed dermatoscopic images from the training set are fed into the model.

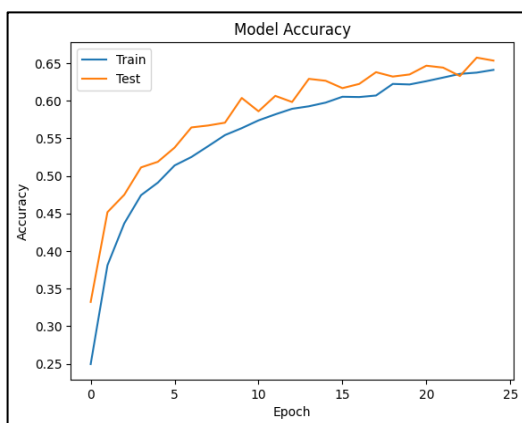
The method uses forward propagation to create predictions using the model's current weights and backward propagation to evaluate the difference between the expected output and the real labels for determining the loss. The optimization technique modifies the model's weights to reduce loss and improve classification accuracy for skin lesions. Regularisation techniques like dropout layers are frequently applied to lessen overfitting and enhance generalization. Until the model reaches a state of convergence, when additional iterations do not significantly increase performance on the training set, the training process iterates over several epochs.

It is essential to look at a different validation set to evaluate the model's capacity to generalize to new data and tune hyperparameters for optimal performance. The model can be tested on a different test set after training and validation to see how well it performs in real-world scenarios when correctly classifying different skin lesions. Regular observation and modification throughout training and thorough testing across several datasets improve the robustness and accuracy of the CNN-Vgg16 skin lesion classification model.

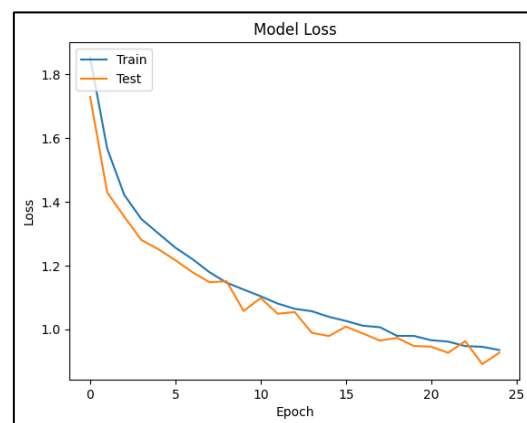
5. RESULTS

This section reports the results of the skin lesion classification technique that was carried out with CNN and Vgg16. Constructing the system on the Google Colab platform was accomplished using Python language.

A. Result of CNN algorithm



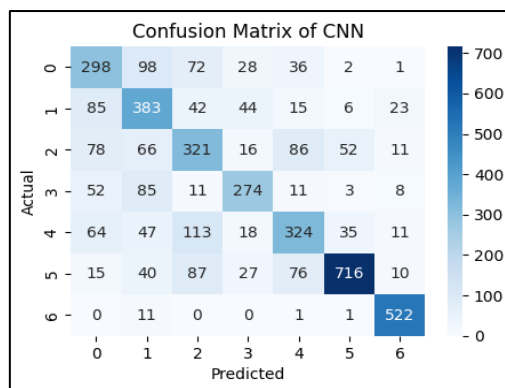
(a)



(b)

Classification Report				
	precision	recall	f1-score	support
akiec	0.50	0.56	0.53	535
bcc	0.52	0.64	0.58	598
bkl	0.50	0.51	0.50	630
df	0.67	0.62	0.64	444
mel	0.59	0.53	0.56	612
nv	0.88	0.74	0.80	971
vasc	0.89	0.98	0.93	535
accuracy			0.66	4325
macro avg	0.65	0.65	0.65	4325
weighted avg	0.67	0.66	0.66	4325

(c)

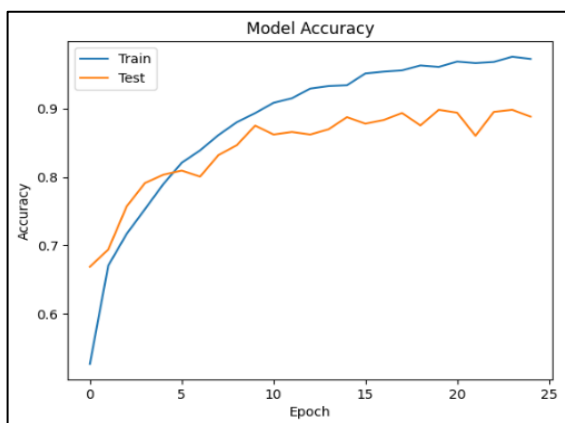


(d)

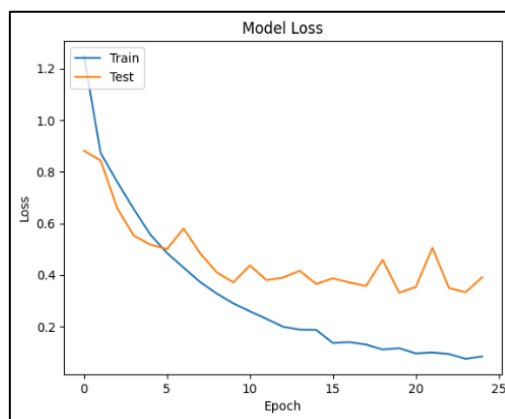
Fig. 4. Result of CNN training progress in terms of (a) Model accuracy, (b) Model loss, (c) Classification report, and (d) Confusion matrix

The results summarize the multi-class skin lesion classification model's performance. Precision, the accuracy of positive predictions, varies from 0.50 to 0.89 across different classes. The recall measures the model's ability to correctly identify instances of a certain class, ranging from 0.56 to 0.98. The F1-score, which fluctuates between 0.50 and 0.93, is a trade-off between recall and precision. The percentage of accurately classified cases in the dataset is 66%, representing the model's overall accuracy. Weighted and macro averages can be useful for evaluating a model's capacity to generalize across several classes. The macro-average metrics provide an average value of 0.65. The results demonstrate the model's advantages, potential areas for improvement, and its capacity to classify various skin lesions accurately.

B. Result of the Vgg16 algorithm



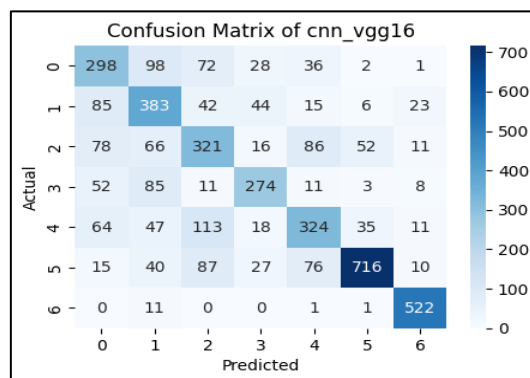
(a)



(b)

Classification Report				
	precision	recall	f1-score	support
akiec	0.94	0.83	0.88	535
bcc	0.93	0.93	0.93	598
bkl	0.77	0.83	0.80	630
df	0.98	0.99	0.99	444
mel	0.78	0.77	0.78	612
nv	0.89	0.90	0.89	971
vasc	0.99	1.00	0.99	535
accuracy			0.89	4325
macro avg	0.90	0.89	0.90	4325
weighted avg	0.89	0.89	0.89	4325

(c)



(d)

Fig. 5. Result of Vgg16 training progress in terms of (a) Model accuracy, (b) Model loss, (c) Classification report, and (d) Confusion matrix

The performance metrics for a skin lesion classification model with seven different classes are summarised by the Vgg16 method. Every row corresponds to a different class, and the columns show the support for each class and the precision, recall, and F1 score. The balanced average of recall and accuracy is known as the F1 score. The accuracy of optimistic forecasts is their correctness. The percentage of real positive instances that are accurately anticipated is known as recall. In the majority of classes, the model has excellent precision and recall. It excels most in classes like "df" and "vasc," where the accuracy and recall values are almost precisely equal to 1.

The percentage of accurately diagnosed cases across all classifications is 89% overall accuracy. A comprehensive view is provided by macro and weighted average measures, where macro averaging treats all classes identically and weighted averaging addresses class differences. The model shows stable performance in various skin lesions with a weighted average F1-score of 89% and a macro-average F1-score of 90%. These metrics explain how well the model can categorize skin lesions into various groups.

Table I presents the CNN and Vgg16 algorithms' comparative study.

TABLE I. COMPARATIVE ANALYSIS OF PERFORMANCE ANALYSIS OF CNN AND VGG16

Algorithm	Precision	Recall	F1 Score	Accuracy
CNN	0.67	0.66	0.66	0.66
Vgg16	0.89	0.89	0.89	0.8883

Table I compares skin lesion categorization performance metrics using two different algorithms: Vgg16 and Convolutional Neural Network (CNN). Each algorithm's accuracy, F1 score, precision, and recall are given. The CNN's accuracy, precision, recall, and F1 scores are all 0.66. With an F1 score of 0.89, the Vgg16 method outperforms the CNN algorithm regarding accuracy, recall, and F1 score.

6. CONCLUSION

The research focused on using CNN and Vgg16 to recognize skin cancer. Our focus was on enhancing early detection, which is crucial for early treatment. This system utilized a substantial dataset,

HAM10000, to evaluate the performance of VGG16 and CNN for automating the categorization of skin lesions. The models demonstrated exceptional sensitivity, specificity, and accuracy, highlighting their potential use as effective tools for dermatologists and other healthcare practitioners. The Vgg16 algorithm surpasses the CNN algorithm in this system.

In the future, beyond the capabilities of Vgg16, investigating the integration of sophisticated architectures and ensemble techniques could be helpful. Expanding the dataset with more diverse sample types could enhance the model's ability to generalize to a wider range of skin illnesses. One interesting strategy to ensure transparency and confidence in model predictions for clinical application is to look into ways to explain and interpret deep learning models in dermatological diagnostics.

7. AUTHOR CONTRIBUTIONS

[Yashwant S. Ingle]: Conceptualization, data curation, methodology, software development, validation, formal analysis, investigation, writing—original draft preparation, writing—review, and editing, visualization, supervision, project administration.

[Dr. N. F. Shaikh]: Conceptualization, data curation, methodology, software development, validation, formal analysis, investigation, writing original draft preparation, writing review, and editing, visualization, supervision, project administration

8. CONFLICT OF INTEREST STATEMENTS

We declare no conflicts of interest regarding the publication of this research manuscript. The research conducted by [Yashwant S. Ingle] was independent and unfunded. No financial or personal relationships with individuals or organizations that could influence the interpretation or presentation of the study's findings exist.

The dataset utilized in this study was sourced from Kaggle, an online platform for datasets, with no associated conflicts of interest regarding its usage or dissemination. All analyses and interpretations presented in this manuscript are conducted with integrity and without bias.

9. ROLE OF FUNDING SOURCE

This research project was conducted without external funding. While funding can often play a crucial role in supporting research endeavors by providing resources and support, this project was carried out independently, without financial assistance.

The absence of external funding allowed for flexibility in the research process, enabling the exploration of various methodologies and techniques without constraints. It also ensured that the research direction and outcomes were driven solely by scientific inquiry and not influenced by funding interests.

While funding can provide valuable resources and support for research projects, the lack of external funding for this project highlights the potential for independent research initiatives to contribute to scientific knowledge and innovation.

10. ETHICS COMMITTEE APPROVAL

This research project involved the utilization of an online dataset obtained from Kaggle for the purpose of skin cancer detection. Since the dataset consisted of anonymized images and did not involve human participants or animals, Ethics Committee approval was not applicable to this study.

The data used in this research were publicly available and did not involve any interactions or interventions with human subjects. Therefore, there were no ethical considerations related to participant welfare, confidentiality, or informed consent.

As such, the research was conducted in accordance with ethical standards for the use of publicly available datasets and adhered to the terms of use specified by Kaggle.

REFERENCES

- [1] Masood, Ammara & Al-Jumaily, Adel. (2013), "Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms", *International journal of biomedical imaging*. 2013. 323268. 10.1155/2013/323268.
- [2] N. Hameed, A. Ruskin, K. Abu Hassan, and M. A. Hossain, "Comprehensive survey on image-based computer aided diagnosis systems for skin cancer," 2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), Chengdu, 2016, pp. 205-214, doi: 10.1109/SKIMA.2016.7916221.
- [3] Ogorzalek, Maciej & Surówka, Grzegorz & Nowak, Leszek & Merkwirth, Christian. (2010). *Computational Intelligence and Image Processing Methods for Applications in Skin Cancer Diagnosis*. *Communications in Computer and Information Science*. 52. 3-20. 10.1007/978-3-642-11721-3_1.
- [4] F. Santos, F. Silva, and P. Georgieva, "Automated Diagnosis of Skin Lesions," 2020 IEEE 10th International Conference on Intelligent Systems (IS), Varna, Bulgaria, 2020, pp. 545-550, doi: 10.1109/IS48319.2020.9200090.
- [5] S. R. Guha and S. M. Rafizul Haque, "Convolutional Neural Network Based Skin Lesion Analysis for Classifying Melanoma," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2019, pp. 1-5, doi:10.1109/STI47673.2019.9067979.
- [6] Mengistu, A. and Dagnachew Melesew Alemayehu. "Computer Vision for Skin Cancer Diagnosis and Recognition using RBF and SOM." (2015).
- [7] Uzma Bano Ansari, Tanuja Sarode, "Skin Cancer Detection Using Image Processing," *International Research Journal of Engineering and Technology (IRJET)*, Volume: 04 Issue: 04 | Apr -2017, pp. 2875-2881
- [8] E. Jana, R. Subban and S. Saraswathi, "Research on Skin Cancer Cell Detection Using Image Processing," 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, India, 2017, pp. 1-8, doi: 10.1109/ICCIC.2017.8524554.
- [9] A. Masood, A. Al-Jumaily and K. Anam, "Self-supervised learning model for skin cancer diagnosis," 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, France, 2015, pp. 1012-1015, doi:10.1109/NER.2015.7146798.
- [10] M, Vijayalakshmi. (2019). Melanoma Skin Cancer Detection using Image Processing and Machine Learning. *International Journal of Trend in Scientific Research and Development*. Volume-3. 780-784. 10.31142/ijtsrd23936
- [11] Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* 5, 180161 doi:10.1038/sdata.2018.161 (2018).