International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Set-Oct-2024, ISSN (Online): 2395-566X

Advanced Skin Cancer Detection using Hybrid CNN Feature Extraction

Mr. S. Sinimoxon Lee, Professor Arpita Das
School of Computer Science Engineering
Vaishnavi Institutes of Technology and Science Bhopal, India

Abstract- Skin cancer is one of the deadliest types of cancer, with a rapidly increasing incidence worldwide. Early detection is crucial to reducing the mortality rate. In this paper, we present an effective computer-aided diagnostic model for accurate skin cancer detection and classification. Our proposed system consists of three primary steps: a) Preprocessing, b) Feature extraction, and c) Classification. During preprocessing, image quality is enhanced through median filtering. In the feature extraction phase, features are extracted from three powerful pretrained CNN models—GoogleNet, AlexNet, and ResNet-101—using transfer learning and are then combined. In the classification stage, the hybrid features are classified using three successful Machine Learning (ML) classifiers: Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN). We validated our model on 3000 images from the MNIST dataset, achieving an accuracy of 96.66%, a precision of 96.5%, a recall of 96.66%, and an F1-score of 96.5%.

Index Terms- Feature extraction, Skin cancer, Convolution Neural Network, Machine Learning

I. INTRODUCTION

Skin cancer is a common type of cancer, as skin cancer affects more than millions of People every year. Cancer can start anywhere in the human body, and if it is not identified and treated in a timely manner, it can be fatal. In paper [1] author proposed An ensemble deep learning model for the classification he aggregated three pre trained CNN models and took classification decision based on weighted output of each CNN classes of skin cancer namely nevus, keratosis and melanoma [2]. Convolutional Capsule Network (Caps Net) was designed for skin cancer classification that can work well with high dimensional hyper spectral images of skin. an intelligent and rapid classification system as proposed in [3] which utilised pre-trained Alex net features and ECOC SVM classifier and achieved satisfactory results. In [4] the skin lesion was segmented using K-mean clustering and extracted the most suitable features by hybridizing color and texture features, then finally classified using SVM. In [5] an CAD system for skin cancer image classification is developed using fine-tuned pertained CNNs and obtained the maximum accuracy of 92.83.IN [6] the author presented customized multilayer Perceptron network and VGG-16 for skin cancer classification and his results proved that VGG-16 gives highest accuracy then other .IN [7] the prediction capability of dermatologist and CNN based artificial intelligent technique and consulted that AI model is provided superior result than dermatologist. Segmentation based skin cancer detection was presented in[18], that used modified Otsu and watershed segmentation algorithm for segmenting cancer region and extracted the CNN features and classified with Back propagation algorithm and SVM.A new method for skin cancer diagnosis developed in [8]which Segmented the Region of interest using mean shift segmentation from which GLCM,GLRLM and moment variants features are extracted and classified by RF-SVM classifier . This paper [9] used statistical and texture GLCM from the segmented images and classified with multiclass SVM. In this paper [10]skin lesions are classified based on, retrained CNNS .and achieved better results. Hybrid features base skin cancer classification was propped in [11]which utilize LBP and GLCM features and finally classified with Artificial Neural Network algorithm. An intelligent region of interest (ROR) based system was developed in [12],which extracted ROI using K-means algorithm, from that the features are extracted classified using pertained CNNS. A classification model with less complex architecture was built in [13]. Which utilize mobile net v2 deep learning model, for extracting the most discrimination features with good classification accuracy.

The main contributions of our work are:

- We proposed a classification framework for classifying skin cancer even with less number of images.
- We ensemble CNN features to create the most effective feature subset.
- We utilized different machine learning classifiers to find the best classifier.
- The robustness of our mode is evaluated using three performance metrics like Accuracy, Precision and Recall.

© 2024 IJSRET

Our paper was organised in the following way: In section 2 we have given the methodology. In section 3 we have discussed our results. Finally, we concluded in section4.

II. PROPOSED METHODOLOGY

The proposed methodology for skin cancer detection and classification is designed to maximize accuracy and efficiency through a structured three-step approach. First, the system begins with preprocessing, where median filtering is applied to the input images to enhance their quality by reducing noise while preserving important features. Next, in the feature extraction stage, we leverage the power of transfer learning by utilizing three well-established pretrained Convolutional Neural Network (CNN) models: GoogleNet, AlexNet, and ResNet-101. Features extracted from these models are then combined to form a robust set of hybrid features that capture various aspects of the input images. Finally, in the classification step, the hybrid feature set is classified using three high-performing Machine Learning (ML) classifiers—Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN)—to ensure precise identification and categorization of skin cancer types. This methodology is validated on a dataset of 3000 images, demonstrating the system's effectiveness in achieving high accuracy and reliability in skin cancer detection. The flow diagram of our proposed model is displayed in Figure 1.

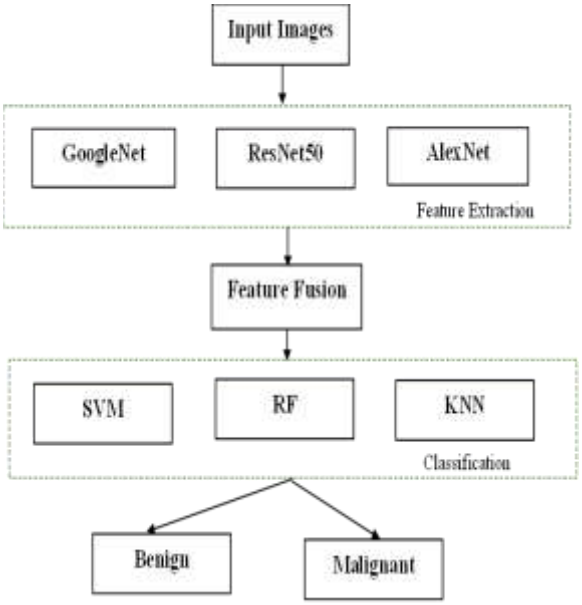

Fig.1. Flow diagram of proposed model

1. Dataset

We have collected 3000 skin cancer images from publicly available MNIST Datasets. The collected Dataset consists of two different classes of skin cancer images called benign and malignant. benign classes have 1500 images and malignant classes have 1500 images.

Fig.2. Sample images of MNIST dataset a) Benign b) Malignant

2. Convolutional Neural Networks

In recent years, significant advancements have been made in natural image classification through the development and publication of various Convolutional Neural Network (CNN) models, such as GoogLeNet[14], AlexNet[15], and ResNet[16], among others. These models have been pivotal in advancing the accuracy and efficiency of image classification tasks. Some of these models, particularly GoogLeNet and ResNet, are provided as pre-trained models, which means they have already been trained on extensive datasets, making them readily usable for new tasks without the need to start training from scratch. Specifically, these models were originally trained on the ImageNet dataset, which consists of around 1.28 million real-world images [25]. ImageNet is a large-scale dataset that contains images from a vast range of categories, enabling the models to learn and generalize features across diverse visual data. By leveraging the power of these pre-trained models, which have already captured complex patterns and features from the ImageNet dataset, we applied them to our task using a technique known as transfer learning. Transfer learning allows us to adapt these pre-trained models to our specific image classification task by fine-tuning them, thereby achieving effective results without the need for extensive retraining. This approach significantly reduces the computational effort and time required to develop a high-performing model for our specific application.

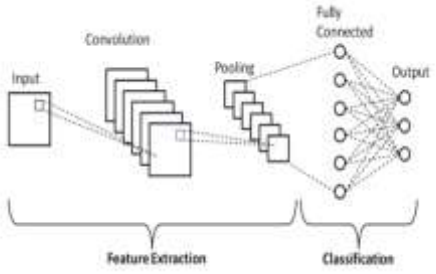

Fig.3. General Diagram of CNN

© 2024 IJSRET

Transfer Learning

Convolutional Neural Networks (CNNs) achieve high diagnostic accuracy in medical image classification by directly extracting features from the images. However, training CNNs from scratch for medical imaging is difficult due to the limited size of available databases. To overcome this challenge, pre-trained CNNs, which have been trained on over a million images for non-medical purposes, are employed to extract relevant cancer information from medical images. In this study, two of the most effective pre-trained networks, DenseNet 201 and Inception ResNet v2, were examined. Among these, DenseNet provided superior results compared to Inception ResNet v2. Consequently, the transfer learning process was demonstrated using the DenseNet 201 network, as depicted in Fig. 4. Moreover, when trained on smaller datasets, the deep connections in these networks help reduce overfitting through a regularization effect [19].

3. Classification

Random Forest [17] is a robust supervised learning algorithm that leverages ensemble learning to boost model accuracy and handle complex tasks. It constructs multiple decision trees during training by creating various subsets of the input data, which helps in capturing different patterns and reducing biases. Each decision tree makes an individual prediction, and the final output of the Random Forest is determined by averaging these predictions (for regression tasks) or using majority voting (for classification tasks). This ensemble approach improves the model's performance and generalization by mitigating overfitting, which is a common issue with single decision trees. By combining the strengths of multiple decision trees, Random Forest enhances predictive accuracy and provides a more reliable and stable model.

III. RESULTS AND DISCUSSIONS

Before our proposed method is trained on MATLAB 2020a. We have collected 3000 images from the MNIST dataset, 80% for training and 20% for testing. The reason for this process is to explore the effectiveness of our model with fewer images. Here we used CNN as only the feature extractor. We extracted the features of three pertained CNNs and integrated to form strong features. Those high dimensional features are reduced using PCA dimensionality reduction technique. Eventually the selected features are trained using four ML classifiers. We also evaluated the individual performance of the four ML Classifiers namely SVM, KNN, RF and Ad boost and the shown in Fig.4-Fig.8. It is observed that the classification performance improved with hybrid features. Among classifiers, the Random Forest Classifier achieved the highest accuracy of 96.66% Precision of 96.5% recall of 96.66%, and f1-score of 96.5%. When comparing the individual performances, the performance of Resent 101is better than Alex net and Google net in Skin cancer image classification.

The confusion matrix of our proposed method is shown in fig.9.

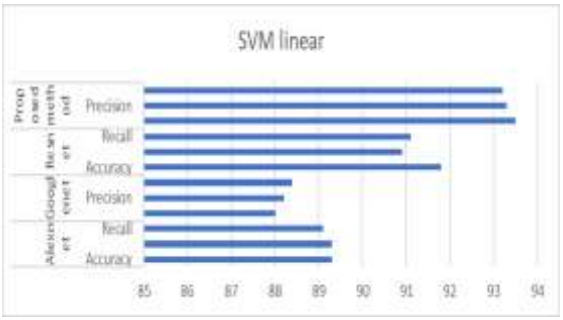

Fig.4. Classification of Hybrid features using SVM with linear kernel

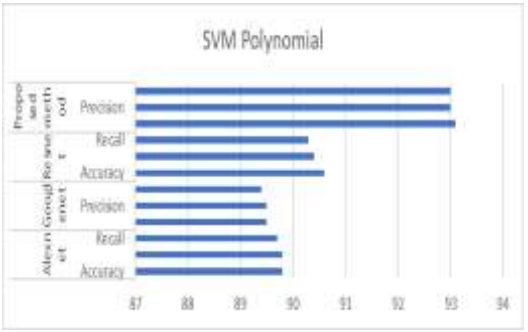

Fig.5. Classification of Hybrid features using SVM with Polynomial kernel.

Fig.6. Classification of Hybrid features using SVM with Radial Basis Function (RBF) kernel

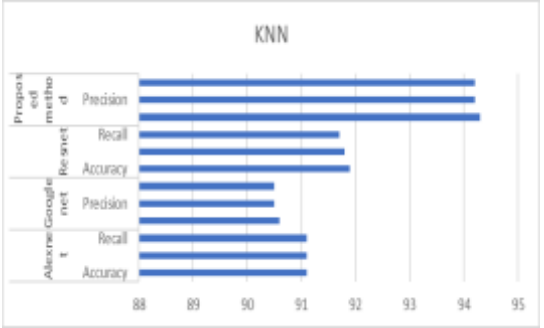

Fig.7. Classification of Hybrid features using KNN classifier

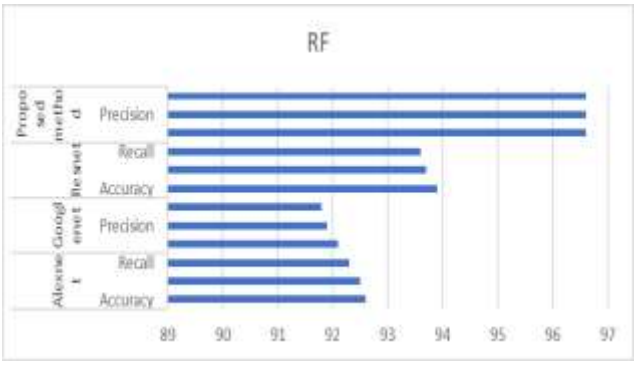

Fig.8. Classification of Hybrid features using Random Forest classifier

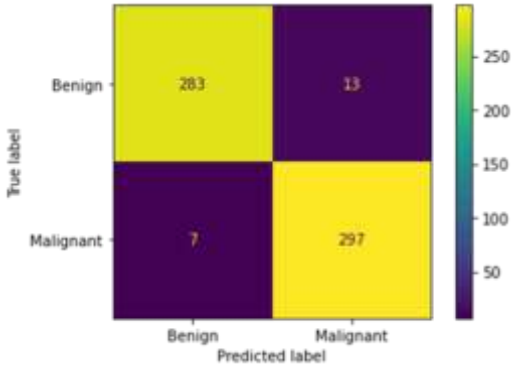

Fig.9.Confusion matrix of proposed model

Data Availability: Data analyzed during the current study are openly available at location https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

Declarations Ethical Approval: This study is the authors' own original work, which has not been previously published elsewhere.

Conflict Of Interest: The authors declare that they have no conflict of interest.

IV. CONCLUSION

This study presents a new method for classifying skin cancer images, highlighting its effectiveness even with a limited dataset. To demonstrate the model's capability in situations where medical images are scarce, we utilized a reduced subset of the MNIST dataset, which, while not medical in nature, serves to illustrate our model's robustness under data constraints. Our approach involves enhancing predictive accuracy through a combination of features extracted using ensemble techniques from several well-known Convolutional Neural Networks, namely ResNet-101, AlexNet, and GoogLeNet. These networks are integrated to provide a comprehensive feature set, which is then used for classification with various machine learning algorithms. Among the classifiers evaluated, the Random Forest classifier delivered the best performance, achieving an accuracy of 96.6%. This demonstrates the effectiveness of our technique in accurately classifying skin cancer images, even when only a limited amount of data is available.

REFERENCES

1. B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," J. Biomed. Inform., vol. 86, pp. 25–32, Oct. 2018, doi: 10.1016/j.jbi.2018.08.006
2. M. Vimal Cruz, A. Namburu, S. Chakkaravarthy, M. Pittendreigh, and S. C. Satapathy, "Skin Cancer Classification using Convolutional Capsule Network (CapsNet)," 2020.
3. R. C. Maron et al., "A benchmark for neural network robustness in skin cancer classification," Eur. J. Cancer, vol. 155, pp. 191–199, Sep. 2021, doi: 10.1016/j.ejca.2021.06.047.
4. M. Q. Khan et al., "Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer," IEEE Access, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2926837.
5. S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan, "A multi-class skin Cancer classification using deep convolutional neural networks," Multimed. Tools Appl., vol. 79, no. 39–40, pp. 28477–28498, Oct. 2020, doi: 10.1007/s11042-020-09388-2.
6. Y. Jusman, I. M. Firdiantika, D. A. Dharmawan, and K. Purwanto, "Performance of multi layer perceptron and deep neural networks in skin cancer classification," in LifeTech 2021 - 2021 IEEE 3rd Global Conference on Life Sciences and Technologies, Mar. 2021, pp. 534–538. doi: 10.1109/LifeTech52111.2021.9391876.
7. A.Hekler et al., "Superior skin cancer classification by the combination of human and artificial intelligence," Eur. J. Cancer, vol. 120, pp. 114–121, Oct. 2019, doi: 10.1016/j.ejca.2019.07.019.
8. A. Murugan, S. A. H. Nair, A. A. P. Preethi, and K. P. S. Kumar, "Diagnosis of skin cancer using machine learning techniques," Microprocess. Microsyst., vol. 81, Mar. 2021, doi: 10.1016/j.micpro.2020.103727.
9. M. Krishna Monika, N. Arun Vignesh, C. Usha Kumari, M. N. V. S. S. Kumar, and E. Laxmi Lydia, "Skin cancer detection and classification using machine learning," in Materials Today: Proceedings, 2020, vol. 33, pp. 4266–4270. doi: 10.1016/j.matpr.2020.07.366.
10. Ravi Manne, Snigdha Kantheti and Sneha Kantheti, "Classification of Skin cancer using deep learning, Convolutional Neural Networks - Opportunities and vulnerabilities- A systematic Review," Int. J. Mod. Trends

© 2024 IJSRET

Sci. Technol., vol. 6, no. 11, pp. 101–108, Nov. 2020, doi: 10.46501/ijmtst061118.
11. F. W. Alsaade, T. H. H. Aldhyani, and M. H. Al-Adhaileh, “Developing a Recognition System for Diagnosing Melanoma Skin Lesions Using Artificial Intelligence Algorithms,” Comput. Math. Methods Med., vol. 2021, 2021, doi: 10.1155/2021/9998379.
12. R. Ashraf et al., “Region-of-Interest Based TransferLearning Assisted Framework for Skin Cancer Detection,” IEEE Access, vol. 8, pp. 147858–147871, 2020, doi: 10.1109/ACCESS.2020.3014701.
13. A.Wibowo, C. A. Hartanto, and P. W. Wirawan, “Android skin cancer detection and classification based on mobilenet v2 model,” Int. J. Adv. Intell. Informatics, vol. 6, no. 2, pp. 135–148, Jul. 2020, doi: 10.26555/ijain.v6i2.492.
14. C. Szegedy et al., “Going deeper with convolutions,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, vol. 07-12-June-2015. doi: 10.1109/CVPR.2015.7298594.
15. A.Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Commun. ACM, vol. 60, no. 6, 2017, doi: 10.1145/3065386.
16. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-December. doi: 10.1109/CVPR.2016.90.
17. A. Murugan, S. A. H. Nair, and K. P. S. Kumar, “Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers,” J. Med. Syst., vol. 43, no. 8, 2019, doi: 10.1007/s10916-019-1400-8.
18. M. A. Ahmed Thaajwer and U. A. Piumi Ishanka, “Melanoma skin cancer detection using image processing and machine learning techniques,” 2020. doi: 10.1109/ICAC51239.2020.9357309.
19. Joseph, J. Sharmila, Abhay Vidyarthi, and Vibhav Prakash Singh. "An improved approach for initial stage detection of laryngeal cancer using effective hybrid features and ensemble learning method." Multimedia Tools and Applications 83.6 (2024): 17897-17919.19.

© 2024 IJSRET