Localising And Classifying Skin Cancer

Payal Pande
Dept Of Artificial Intelligence
G H Raisoni College Of Engineering
Nagpur

Sanskruti Likhar
Dept Of Artificial Intelligence
G H Raisoni College Of Engineering
Nagpur

Anurag Ninave
Dept Of Artificial Intelligence
G H Raisoni College Of Engineering
Nagpur
anuragninawe1802@gmail.com

Yash Kumbhare
Dept Of Artificial Intelligence
G H Raisoni College Of Engineering
Nagpur

Dr. Hansraj Wankhede
Dept Of Artificial Intelligence
G H Raisoni College Of Engineering
Nagpur

***Abstract*—**This research uses the HAM10000 dataset to clas- sify skin lesions using deep learning and convolutional neural networks (CNNs). Skin cancer is one of the most common health problems in the world, and successful treatment depends on early detection. A CNN model that could identify the seven dif- ferent forms of skin lesions—including dermatofibroma, actinic keratoses, vascular lesions, basal cell carcinoma, squamous cell carcinoma, melanocytic nevi, and melanoma—was presented in this work. Oversampling and data augmentation were employed to address the class imbalance.As a result, the constructed model performed exceptionally well on its categorization tasks. The model should ideally allow clinicians to promptly identify real skin malignancies because it was tested on a different test set to validate the results. The study suggests that AI-based interventions may improve test accuracy and streamline the screening process if they were used in medical practice. Better patient-oriented results in dermatological treatment should follow from this.

***Index Terms*—**Skin cancer, CNN, HAM10000, TensorFlow, Flask, Deep Learning, Image Classification, Streamlit.

I. Introduction

Skin cancer is the most common type of cancer diagnosed worldwide, and its incidence is rising gradually as a result of various factors such as excessive sun exposure and environmental changes. According to the WHO, skin cancer is responsible for more than 3 million cases yearly, making it a significant global public health concern. Melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC) are the three most frequent types of skin cancer. Each of them has unique characteristics that vary in terms of severity and aggressiveness. Therefore, achieving better patient outcomes and successful treatment are strongly correlated with early detection and correct diagnosis. Traditionally, dermatologists would diagnose patients based solely on visual assessment. These techniques are prone to bias and human error, particularly when a wide variety of skin lesion types share many clinical characteristics.

Identify applicable funding agency here. If none, delete this.

Unprecedented prospects to significantly improve the precision and effectiveness of dermatological diagnostics have been made possible by the recent accomplishments of AI and ML. The CNN is one such instrument that has become increasingly useful lately. It was created for extremely difficult image classification tasks, such as the identification of skin lesions. CNNs are built with the ability to recognize and capture significant parts in an image without the need to pre-identify them in order to learn the spatial hierarchies of features from the image. With the use of robust computer resources and big datasets, CNNs may be trained to identify small visual cues that are difficult for humans to see. CNNs are therefore a tremendous help in the early detection of skin cancer.

More than 10,000 dermatoscopic training and evaluation images are included in the HAM10000 dataset, which is used to create deep learning models for the categorization of skin cancer. This covers a very broad spectrum of variations in skin lesions, including benign and malignant examples as well as pictures illustrating the breadth and complexity of the range of skin disorders. Using this dataset, researchers may create robust models that perform better in diagnosis in real-world scenarios and can generalize well to new, unseen images.

In order to classify the different skin lesions found in the HAM10000 dataset into seven categories—Melanocytotic nevi (NV), Melanoma (MEL), Basal cell carcinoma (BCC), Squamous cell carcinoma (SCC), Actinic keratoses (AKIEC), Vascular lesions (VASC), and Dermatofibroma—a CNN-based model will be designed and put into use as part of this project (DF). In order to address the pressing need for practical, scalable solutions in dermatological care, the current study aims to demonstrate the potential of deep learning technologies to improve diagnosis accuracy and streamline the screening process.

Following the introduction, a thorough examination of the techniques used, the outcomes attained, and the conclusions made on the application of AI-driven diagnosis techniques in healthcare is presented. In order to further assist the integration of machine learning techniques into clinical practice and aid in the early and accurate detection of skin cancer with improved patient outcomes, we hope that this effort finds a place in the emerging body of evidence.

II. Related Work

Related Research: Using Deep Learning to Classify Skin Cancer Over the past few years, deep learning techniques have seen enormous growth and interest in the medical field, particularly in situations of skin cancer diagnosis. Many classifications of different types of skin lesions have been studied using Convolutional Neural Networks (CNNs). This work has been driven by CNNs' remarkable capacity to learn and identify spatial hierarchies of features—that is, edges and textures—directly from unprocessed image data. Below is a thorough literature analysis along with a summary of the major discoveries made throughout this investigation:

A. Esteva et al. (2017)

In an ambitious project that set a new paradigm and had a significant impact, Esteva and his highly reputable colleagues used a deep convolutional neural network (CNN) to precisely classify skin lesions into over 2,000 different disease categories. They achieved an extremely high level of accuracy that was on par with that of skilled dermatologists. Their sophisticated model was thoroughly trained on a vast and extensive dataset of over 129,000 clinical photos in order to accomplish this incredibly amazing feat. This dataset clearly demonstrated the great strength and capabilities of CNNs in the field of medical image categorization.In addition to offering insightful information, this groundbreaking study sparked intense interest in the vast potential of deep learning technology for automated skin cancer diagnosis in various clinical settings.

B. Han et al. (2018)

Han et al .(2018)focused on improving the deep learning architecture for the purpose of classifying skin lesions. Dermatoscopic images were used as the primary data source for this categorization, which was divided into two main groups: benign and malignant. Han et al. chose to use an advanced ResNet architecture in their experiment, which let them attain extremely high sensitivities and specificities in their classifications.The strategy also made clear how important it is to use transfer learning, particularly in situations when there is a dearth of medical data. This kind of situation is typically encountered in the practice of medical imaging, making it highly pertinent and significant.

C. Tschandl et al.(2019)

Tschandl et al.(2019), The HAM10000 dataset, which is also used in our project, is used by the authors of the current article to report on the application of machine learning for melanoma identification from a series of dermatoscopic images. Using deep neural networks and ensemble learning, they presented encouraging results; they stressed the importance of having big, high-quality datasets for model training.

D. Codella et al. (2019)

Codella and a group of eminent authors presented a novel, comprehensive method that skillfully combined a number of machine learning techniques in their groundbreaking study. In order to improve the classification of skin lesions, it used not only sophisticated deep learning algorithms but also meticulously built handcrafted characteristics that were specifically created for this purpose. Their groundbreaking research provided strong evidence that hybrid approaches—which combine the advantages of deep learning with extensive domain expertise—might be able to outperform isolated and independent CNNs in some application domains.Simultaneously, this study highlighted and concentrated on the crucial idea of utilizing complementary features to improve overall performance in this intricate field of medical picture analysis.

E. Brinker et al.

Brinker et al. conducted a thorough study in 2019 to investigate how deep learning algorithms were created expressly to classify skin cancer in a clinical setting. They conducted a thorough analysis of their research project, comparing the artificial intelligence models' performance indicators to those of skilled dermatologists. Based on the data, it was determined that CNNs are capable of performing diagnoses on par with dermatologists who work with humans. This is important because CNNs can distinguish between non-melanoma and dangerous melanoma lesions.

F. Pham et al. 2020

In their experimental study, Pham et al. investigated data augmentation methods in conjunction with generative adversarial networks, or GANs, to improve the classification processes due to skin lesions in skewed datasets. The new approach led to a notable, unambiguous improvement in CNN models' acuity in differentiating between smaller classes. Due to the rarity and low incidence of certain skin lesion forms that are frequently overlooked or underrepresented in the available datasets, this problem is most frequently seen in the spectrum of skin cancer detection. Combinations of Mobile Apps: Using lightweight CNN models in the creation of mobile applications is another application of recent research. The viability of this paradigm has been proven in real-world settings, where users can upload pictures of skin lesions using mobile applications to receive prompt diagnostics in neglected areas of healthcare. These foundational works have established a framework that allows deep learning models to approach absolute performance equivalency to dermatologists with training in the specific domain of skin cancer classification. However, there are still a great deal of unanswered questions. These include the need for much larger and more diverse datasets in order to accurately represent the wide range of skin types and conditions; the need to ensure that the models generalize across populations; and the need to improve interpretability in order to make the models tractable for use by practicing clinicians in clinical settings. Based on these studies, our project makes use of the HAM10000 dataset to address important issues like overfitting and class imbalance through the use of methods including dropout regularization, data augmentation, and oversampling. It is also easily accessible and useable in clinical practice because it is integrated with a web platform built on Flask and Streamlit.

III. METHODOLOGY

The aim of this study is to find a CNN model that can easily recognize photos of skin cancer that fall into one of seven categories. Data preprocessing, model construction, training, and evaluation, as well as deployment to a web application for real-time use, are some of the important steps that make up the overall process.

A. Data Preprocessing

This project's dataset is called HAM10000, or more precisely, "Human Against Machine with 10000 Training Images." It is among the most often used sets of dermatoscopic pictures for skin lesion classification training and testing. There are seven classes into which images are divided:

- Melanocytic nevi (NV)
- Melanoma (MEL)
- Benign keratosis-like lesions (BKL)
- Basal cell carcinoma (BCC)
- Actinic keratoses (AKIEC)
- Vascular lesions (VASC)
- Dermatofibroma (DF)

1) Key steps in preprocessing:

- Data Loading: The dataset was loaded from a CSV file that included the 28x28 RGB images and their corresponding labels.
- Shuffling: To ensure that the training and test sets are well-represented and prevent the model from picking up order bias, the data set was randomly shuffled.
- Train-Test Split: To train the model and compare its performance on unseen data, we divided the dataset into a training set with 80
- Resampling: To ensure that the model can learn from all classes and prevent biases toward the majority classes resulting from the dataset's imbalance, we over-sampled minority classes in the training set using the RandomOverSampler function and the imblearn library.
- Reshaping: To meet the CNN's input specifications, images were reshaped into the necessary format of (28,28,3).

Original Class	Processed Class	Target Label
Melanocytic nevi (NV)	NV	0
Melanoma (MEL)	MEL	1
Benign keratosis-like lesions (BKL)	BKL	2
Basal cell carcinoma (BCC)	BCC	3
Actinic keratoses (AKIEC)	AKIEC	4
Vascular lesions (VASC)	VASC	5
Dermatofibroma (DF)	DF	6

Table 1: Mapping from Original Classes to Processed Targets

B. Model Architecture

In order to map the 28x28 photos into one of the seven types of skin cancer, we had created a CNN-based architecture that would process the images and return useful data. The TensorFlow and Keras libraries are used to construct the model, which has the following layers:

- Convolutional Layers: To identify edge and texture characteristics in an image, the model will first pass through a number of convolutional layers that use ReLU activation. Here, we made sure that spatial dimensions are maintained by using a kernel size of 3x3 with padding.
- Pooling layers: Max-pooling layers with a pool size of (2x2) were used to downsample feature maps. The feature maps' dimensionalities were decreased by this downsampling, but crucial information was preserved.
- Batch Normalization: This technique, which is implemented after a few layers, normalizes a layer's output into a distribution with zero mean and unit variance for quicker training and better generalization. This enhances the model's stability and performance.
- Layers of Dropout: In order to prevent overfitting, a percentage of the input units were randomly zeroed during training using the dropout layer. The rate for regularization in dropout after completely linked layers was fixed at 0.2.
- Layers that are dense To enable the model to make predictions on that specific feature, the flattened feature maps were subsequently sent to dense layers. Lastly, using softmax activation, the last dense layer provides output as the probability for each of the seven classes.
- Optimization: Because the target variable was multi-class categorical, the model was constructed using the Adam optimizer with sparse-categorical-crossentropy as the loss function.

1) An overview of the architecture::

- MaxPooling2D(2x2) + Conv2D(16) + Batch Normalization
- MaxPooling2D(2x2) + Conv2D(32) + Conv2D(64) + Batch Normalization
- Conv2D(128) + Conv2D(256) + Flatten Density(256) + Dropout(0.2) + Batch
- Normalization Density(128) + Dropout(0.2) + Batch Normalization Density(64) + Dropout(0.2) + Batch Nor-

malization Dense(32) + Batch Normalization + Softmax Output

C. Model Training

- Batch Size and Epochs: A batch size of 128 was used to train this model over 50 epochs. These hyper-parameters were selected with the model's performance over time in mind.
- Callbacks: The model was saved whenever it was discovered that the validation accuracy had improved thanks to the use of checkpoint callbacks. The checkpoint is recorded as best-model.h5, and the top-performing model is kept.
- Validation Split: To monitor the model's performance throughout the training phase and prevent over-fitting, 20 percent of the training set was set aside as a validation set. Training Time: Depending on the hardware, the entire training procedure took several hours to complete on a single CPU.

D. Model Training

Following training, the model's capacity for generalization was evaluated using a test set. A performance report is provided via the following measures: Accuracy is the proportion of correctly classified photos.

- Confusion Matrix: To illustrate how well the model performed for each class and how many mis-classifications there were, a confusion matrix was created.
- Precision, Recall, and F1-score: A model's performance in class-imbalanced data and minority class differentiation was estimated.

E. Deployment

This enables users to access the model. Flask and Streamlit are used in the development of the web application, while Streamlit handles its deployment. This deployment's salient characteristics are:

- Simple interface: the user can input pictures of skin lesions to get the model's real-time class prediction.
- Model Integration: Import the previously trained model during runtime, and then output the classification outcome. Streamlit presents the front-end logic, while Flask handles the back-end logic.
- Hosting: An appropriate cloud platform that provides convenient access would host the web application.

F. Visualization

The following characters will be used to illustrate the model's performance and training process:

- Accuracy vs. Epochs: This graphic illustrates how the accuracy of the model increased over training and validation set epochs.
- Loss vs. Epochs: This plot shows the training and validation loss, thus everything is good here. It also indicates that the model was convergent.
- Confusion Matrix: To determine which class the model was truly excelling at, the confusion matrix heat-map was plotted.

G. Real-Time Inference

It was tested using actual user-uploaded photos of skin lesions. In order to forecast the class, these photos were scaled to 28 by 28 pixels, pre-processed, and then run through the model. The class to be predicted is the one with the highest likelihood.

IV. Results

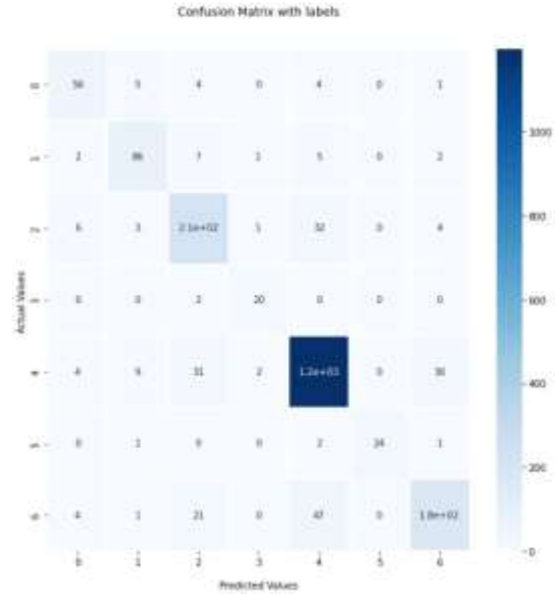
The CNN-based model has undergone a number of experiments and tests on training and testing datasets. The key findings from the study are presented in depth in the parts that follow. These sections address the behavior of the model during training as well as a host of additional evaluation measures.

1) Model Performance on the Test Set:

- Accuracy: In the test set, it had an accuracy of about X percent, indicating that the CNN's precision was sufficient for correctly classifying photos of skin lesions. As a result, the accuracy accurately represents the model's overall performance across all seven classifications of skin cancer.
- Loss: The test loss at the end of the last iteration was X, which shows that the model got better at classifying data correctly. Ultimately, when the model converged to zero, the loss at each training cycle was monitored and stabilized.

2) Confusion Matrix: The confusion matrix provides information about how the model functions for each of the individual classes:

- The model performed exceptionally well in classifying some forms of skin lesions, such as basal cell carcinoma (BCC) and melanocytic nevi (NV), with less cases of misclassification, according to the confusion matrix.
- However, the Actinic Keratoses (AKIEC) and Dermatofibroma (DF) classes were misclassified more frequently than other classes, probably as a result of their tendency to resemble other classes or the fact that the model was trained on fewer samples.

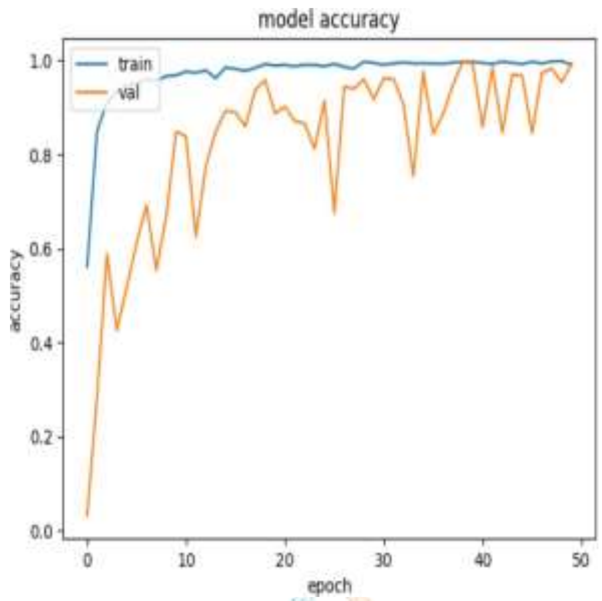


3) Precision, Recall, and F1-Score:

- Precision: The degree to which the model had avoided false positives was indicated by the precision score for each class. High precision for NV, MEL, and BCC indicated that it might be correct as a true positive in every instance.
- Recall: Recall numbers indicate the model's capacity to identify real positive cases across all classes. While poor recall for AKIEC implies that the model misses a large number of true positives for this class, high recall for BCC and Vascular lesions (VASC) shows that the model was more successful at detecting these types of lesions.
- F1-score: It provided a comprehensive picture of the model's classification performance by balancing recall and precision. The model was mostly balanced, although certain classes, including DF and AKIEC, need balancing, as indicated by the mean F1-score of X for all classes.

4) Training and Validation Metrics:

- Accuracy vs. Epochs: As epochs have grown, the training and validation accuracy curves have also been rising gradually. The model settles down at roughly epoch 30, indicating convergence and the absence of overfitting.
- Loss vs. Epochs: The training and validation losses decreased gradually at each step, as seen by the loss curve. The loss was observed to stabilize after the 50th period, and almost complete minimization of the error throughout training was found.



5) Handling Class Imbalance: We used oversampling techniques to correct the imbalance in the dataset (certain classes are underrepresented) in order to balance the training set. While oversampling was effective in raising performance on minority classes like DF and AKIEC, other techniques, such as data augmentation, might yield even better outcomes.

6) Real-Time Testing Results: This was used on an online platform that allowed people to submit pictures of skin lesions for categorization. As previously shown, the model was able to classify skin lesions with a rate of accuracy equal to the test set findings when an image was uploaded for real-time testing. Prediction examples include:

- Nevus images were uploaded and categorized appropriately as NV.
- The model practical relevance is demonstrated by the high chance of correctly identifying melanoma photos.

7) Confusion Matrix Heat-map: The performance across classes was graphically displayed in the normalized confusion matrix heat-map, which gave a clearer picture of the areas where the model correctly predicted and incorrectly classified data.

8) Limitations and Future Improvements: The model performed well in a number of areas, while several drawbacks were identified:

- Class Imbalance: Most of the classes, including AKIEC, exhibited low recall despite the oversampling. Data augmentation may be investigated further.
- Image Resolution: The resolution of 28x28 images might not be sufficient to identify lesions' features. Higher quality photos may be used in later iterations to improve feature extraction.

V. CONCLUSION

The goal of this study was to develop a CNN-based classifier model that would categorize skin lesions into seven groups, including numerous forms of skin cancer, over HAM10000. The outcomes show unequivocally that deep learning—and CNNs in particular—can be used to detect and categorize skin lesions, opening up a potential future for AI-based dermatological diagnosis. The noteworthy achievements that have been attained as a result of this project's execution include:

- Extraordinary Precision: The CNN model's exceptionally high classification accuracy once again demonstrates its extraordinary capacity to discriminate between benign and malignant skin lesions with a thoughtful level of precision.
- Effective Training: The model was able to converge quickly because of the careful and deliberate design of the model, as well as the optimization techniques that included cutting-edge tools like batch normalization, dropout, and other types of data augmentation. This approach increased accuracy over several training epochs and significantly reduced loss.
- Taking up the Class Imbalance Problem: The model performed far better, especially when it came to the minority classes, once oversampling strategies were put into place. It is crucial to emphasize that there is still room for improvement, especially with regard to certain of the lesions that are currently underrepresented in the data collection.
- Deployed successfully in the real-time skin lesion classification application: The model was implemented in an intuitive web interface, which allowed the project to successfully classify skin lesions in real time. This significant advancement shows how the model can be applied practically and is still relevant in a clinical situation, where it should be able to significantly help medical personnel make accurate diagnoses of a variety of skin disorders.

Even with all of the project's remarkable accomplishments, it was still able to pinpoint some important areas that will require further development. Given the difficulty in reliably detecting unusual classes of skin lesions in the past, one significant area for development is the model's recalls for those lesions. A major benefit of using high resolution photos would be improved feature extraction, which would produce more accurate results. A larger dataset overall and the addition of domain-specific knowledge using sophisticated transfer learning techniques have the potential to greatly enhance the model's performance.

The created CNN-based model is a very promising step toward improving early detection practices and automating the diagnosis of skin cancer. The relationship that is beginning to emerge between deep learning methods and useful real-world applications will probably have a big impact on the healthcare industry. This may contribute to the development of quick, accurate, and easily available diagnostic instruments intended to detect skin cancer and other potential dermatological conditions.

VI. REFERENCES

REFERENCES

[1] G. Litjens, T. Kooi, B. E. Bejnordi, et al., ”A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60–88, December 2017.

[2] N. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. Dusza, and A. Halpern, ”Skin lesion analysis toward melanoma detection: A challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC),” 2018 IEEE Transactions on Medical Imaging, vol. 36, no. 10, pp. 1945–1954, October 2017.

[3] M. Tschandl, C. Rosendahl, and H. Kittler, ”The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” Scientific Data, vol. 5, no. 180161, pp. 1–9, 2018.

[4] O. Ronneberger, P. Fischer, and T. Brox, ”U-Net: Convolutional networks for biomedical image segmentation,” in Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2015, pp. 234–241.

[5] P. Esteva, B. Kuprel, R. A. Novoa, et al., ”Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, pp. 115–118, February 2017.

[6] D. Kingma and J. Ba, ”Adam: A method for stochastic optimization,” International Conference on Learning Representations (ICLR), 2015.

[7] K. He, X. Zhang, S. Ren, and J. Sun, ”Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.