IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Skin Cancer Lesion Classification Using Machine Learning

Satyavarapu Varun
Department of Information Technology,
B V Raju Institute of Technology
Affiliated to JNTUH
Vishnupur, Narsapur, Medak,
Telangana State, India

Abdul Malik Khudus Shaik
Department of Information Technology,
B V Raju Institute of Technology
Affiliated to JNTUH
Vishnupur, Narsapur , Medak,
Telangana State, India

Vijaya Katturi, Assistant Professor
Department of Information Technology,
B V Raju Institute of Technology
Affiliated to JNTUH
Vishnupur, Narsapur , Medak,
Telangana State, India

Abstract: The main motive of our image classifier is to classify images of skin cancer, such as basal cell carcinoma, squamous cell carcinoma, melanoma and so on, using Machine Learning algorithms and Convolutional Neural Network models. This model can recognize images based on previously trained images and co- relate the image to the closest form of skin cancer. Classification is a process of categorizing a given set of data into classes, It can be performed on both structured or unstructured data. Image classification is a widely used method for categorizing things, and one area where it's particularly popular is in identifying different types of skin cancer. Skin cancer is a very serious form of cancer that is causing more deaths as people don't always know the symptoms or how to prevent it. The severity of the cancer depends on the type of skin cancer. This research project uses a dataset called HAM10000, which contains over 10,000 images. When creating a model to classify these images, using data augmentation helps the model learn more distinctive characteristics and features than not using data augmentation.

Index Terms – Classification, Skin Cancer, AutoKeras, Data Augmentation.

1. INTRODUCTION

Skin cancer is a type of cancer that starts in the skin cells. The most common types of skin cancer are basal cell carcinoma, squamous cell carcinoma, and melanoma. Basal cell carcinoma and squamous cell carcinoma are usually less dangerous and easier to treat, while melanoma can be more aggressive and deadly if not caught early. Skin cancer is usually caused by exposure to ultraviolet (UV) radiation from the sun or tanning beds. People with fair skin, light- coloured eyes, and a history of sunburns or excessive sun exposure are at a higher risk of developing skin cancer.
Symptoms of skin cancer can include changes in the appearance of moles or other skin growths, such as asymmetry, irregular borders, changes in colour, or an increase in size.

It's important to perform regular skin self-exams and to see a doctor if you notice any changes in your skin. The best way to prevent skin cancer is to protect your skin from the sun by wearing protective clothing, staying in the shade, and applying sunscreen with an SPF of 30 or higher. If skin cancer is detected early, it is highly treatable, but if it spreads to other parts of the body, it can be much more difficult to treat.

You can reduce your risk of skin cancer by limiting or avoiding exposure to ultraviolet (UV) radiation. Checking your skin for suspicious changes can help detect skin cancer at its earliest stages. Early detection of skin cancer gives you the greatest chance for successful skin cancer treatment.

Skin cancer is of many types, and one of the types being more rare and contagious than other, makes the need of proper classification of skin cancer necessary and prominent

2. RELATED WORK

Various machine learning models have been developed to detect skin cancer by analyzing images of skin lesions. One common model is the convolutional neural network (CNN), which is capable of learning patterns in images. CNNs have proven to be effective at identifying skin lesions and detecting skin cancer.

One example of a CNN model used for skin cancer detection is the Inception-v4 model, which was trained on a vast dataset of skin lesion images. The model applies multiple layers of convolutional filters to extract features from the images, followed by pooling

and fully connected layers to predict whether a lesion is cancerous or not. The Inception-v4 model achieved a diagnostic accuracy of 91% on a skin lesion image dataset in a study published in the journal Nature.

Other machine learning models such as support vector machines (SVMs), random forests, and decision trees have also been utilized for skin cancer detection. These models are often combined with image processing techniques such as colour normalization and texture analysis to enhance prediction accuracy. Although machine learning models show potential for improving skin cancer detection accuracy and efficiency, further research is necessary to validate their effectiveness in clinical settings

3. METHODOLOGY

3.1 Visualizing and analysing data:

Visualizing data is very essential before proceeding with training the model. Clear idea about the data and distribution of the different data categories is very essential. Visualization helps in through analyzation of data with respect to categorical distribution.

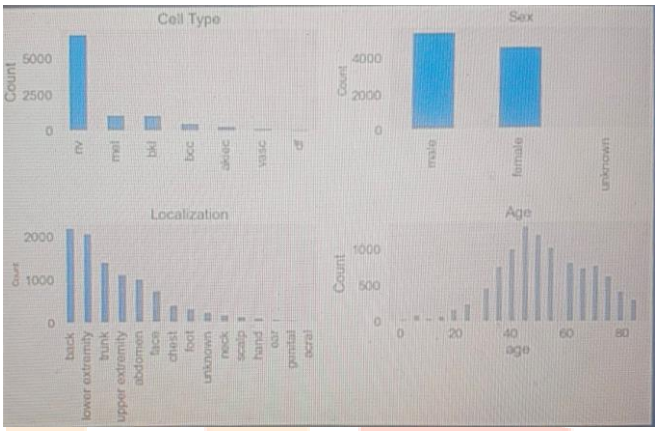

Figure 1. Visualized representation of distribution of data in dataset.

3.2 Resampling and Balancing Data:

The HAM-10000 dataset contains more than 10,000 pictures of skin cancer lesions gathered from various sources, resulting in an uneven distribution of images across categories. To ensure comprehensive training of the model, we can balance the data by resampling each category, which involves augmentation.

3.3 AutoKeras For Model Generation:

AutoKeras is a tool for automated machine learning (AutoML) that simplifies the process of creating neural networks by automating the search for the most effective architecture. To achieve this, AutoKeras uses neural architecture search (NAS) to explore a vast range of potential network architectures and identify the one that performs best for the given task.

The NAS process is guided by a set of heuristics and is based on a validation set. The user only needs to provide the input data and output labels, and AutoKeras takes care of the rest. Once the optimal architecture is identified, AutoKeras trains the model and provides the user with the best model for the task at hand. AutoKeras is versatile, supporting various tasks such as image classification, text classification, and structured data regression. AutoKeras helps minimize the need for human intervention and expertise in neural network design, making it a valuable and accessible tool for researchers and developers.

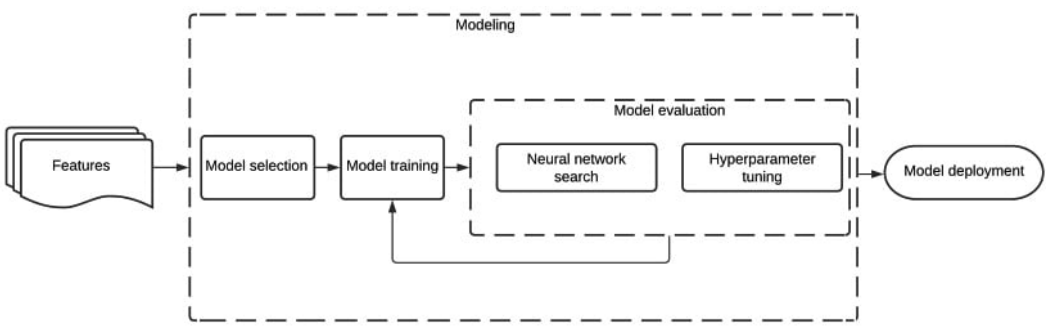

Figure 2. Workflow of Auto Keras

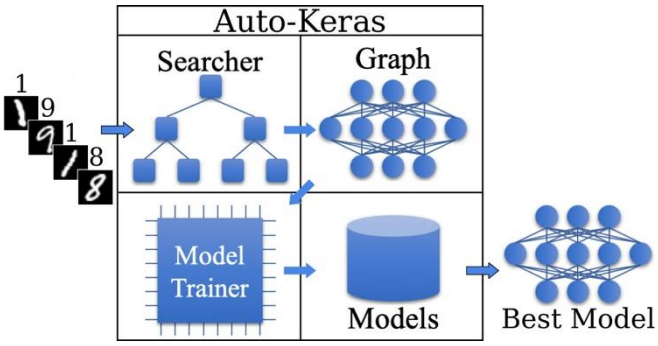

Figure 3. Generalized working of AutoKeras

4. MODULES

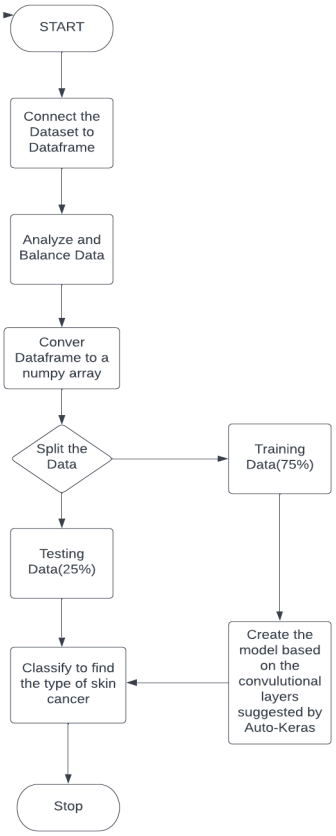

FIGURE 4. WORKFLOW OF OUR MODEL

4.1 Importing the needed Libraries :

To begin with, we need to bring in the essential components required for the upcoming section of our program. NumPy, pandas, matplotlib, and seaborn are the dependencies employed in this project.

4.2 Data Collection & Processing:

- The dataset being used has to be downloaded, in this case "HAM10000" dataset, and connected to our environment, so that code can access necessary data and perform further actions.
- As the downloaded dataset is in a CSV format, we will use PANDAS to read it. Pandas comes with a "read.csv()" method that we can use to read the said CSV file.

4.3 Selection of Attributes & Model:

- The process of attribute selection for a prediction system involves choosing the most relevant attributes, such as age, gender, and location of the patient, in order to improve the efficiency of the system.
- Selection of the appropriate model, is the most important step in our process. We use AutoKeras , by feeding a small subset of our dataset to find the appropriate model.

4.4 Splitting X and Y into training and testing data:

- The splitting of the data is done into four variables, viz., X_train, Y_train, X_test, Y_test.
- ▪ X_train: contains a random set of values from variable 'X, here it is the image of lesion
- ▪ Y_train: contains the output (Label) of the corresponding value of X_train.
- ▪ X_test: contains a random set of values from variable ' X ', excluding the ones
- fromX_train(as they are already taken).
- ▪ Y_test: contains the output () of the corresponding value of X_test.

4.5 Model Training based on the model suggested by auto keras and Result:

- On using AutoKeras with a much smaller subset of our data, the convolutional layers suggested were 256, 128 and 64 to which we have to add a dense layer.
- Once our model is trained, we can further perform classification and showcase our results. We can use various plots and confusion matrix to show our results. Finally, we also show the overall accuracy of our model.
- The accuracy can be increased with help of proper software, and higher accuracy can be achieved.
- On using 100 epochs and 25 max trials to generate the model, The resulting accuracy achieved was 80.7%.

5.Conclusion:

- The average accuracy obtained in the classification of Skin cancer lesion is around 80.7% by using CNN, which can further be increased with help of better hardware.
- This model will provide precise and error-free Predicted data
- Hence, with forthcoming of amazing computational methodologies, and more durable machine learning and deep learning models, under guidance of train professionals advancements can be made to use ML in Health Care industry.

References:

[1] M. R. Foundation. (2016). Melanoma is the Deadliest Form of Skin Cancer. [Online]. Available: https://melanoma.org/

[2] Z. Waheed, A. Waheed, M. Zafar, and F. Riaz, ''An efficient machine learning approach for the detection of melanoma using dermoscopic images,'' in Proc. Int. Conf. Commun., Comput. Digit. Syst. (C-CODE), Mar. 2017, pp. 316–319.

[3] M. Sattar and A. Majid, ''Lung cancer classification models using discriminant information of mutated genes in protein amino acids sequences,'' Arabian J. Sci. Eng., vol. 44, no. 4, pp. 3197–3211, Apr. 2019.

[4] Majid, S. Ali, M. Iqbal, and N. Kausar, ''Prediction of human breast and colon cancers from imbalanced data using nearest neighbor and support vector machines,'' Comput. Methods Programs Biomed., vol. 113, no. 3, pp. 792–808, Mar. 2014.

[5] R. Seeja and A. Suresh, ''Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM),'' Asian Pacific J. Cancer Prevention, vol. 20, no. 5, p. 1555, Feb. 2019.