Prediction of Skin Cancer Using CNN

Shubhada Labde
K.J.Somaiya Institute of Engineering &
Information Technology,
Mumbai, India
shubhada.l@somaiya.edu

Nisha Vanjari
K.J.Somaiya Institute of Engineering &
Information Technology,
Mumbai, India
nvanjari@somaiya.edu

Abstract – **Skin Cancer is a dangerous disease which occurs in almost all age groups especially at an older age. It is due to the abnormal growth or rapid growth in the epidermal layer of the skin, which leads to the growth of tumor. Cancer need to be differentiated from other skin diseases as it occurs on in tissue levels where as other diseases affects to only upper layer of epidemics. As the treatment and detection is at high cost, technology can help predict the cancer at early stage, thereby reducing the detection of cancer's cost athletes and also can help save time. A fully accurate machine to detect cancer stages is difficult to prepare but at least a probability can be given and thus the medical test can further be performed if the accuracy of the machine is not up to mark. So we have proposed a CNN(Convolutional Neural Network) which can predict and classify skin cancer with a better accuracy as compared to normal Support Vector Machine.**

Keywords: CNN, skin cancer prediction, HAM1000 Dataset

I. INTRODUCTION

Skin cancer can be detected using conventional methods and tests but technology can to help it detect to some extent. In the American nation, the majority of people have been suffering from Skin Cancer and the root cause is detection at an early stage. Cancer can be caused due to many reasons such as the harmful rays emitted by the sun, unhealthy eating habits, excessive use of drugs and alcohol, use of plastics for hot storages of food items can too be a cause of cancer [3]. So, technology can help cure as well as detect skin cancer. [11] The parameters used for detecting skin cancer can be categorized as ABCDES (the ABCDEs (Asymmetrical shape, Border irregularities, Color, Diameter, and Evolution) . HAM10000 (Human against Machine) is an open-source database of skin cancer images that is categorized into seven classes of skin cancer stages. The dataset is comprised of 10000 images which is real and genuine and made by using expert and scientists.[3] According to the American Cancer Society (ACS), there were about 5.4 million new cases of skin cancer in the year 2017. The number of new cases for skin cancer exceeds the combined total of new cases for prostate cancer, breast cancer, lung cancer, and colorectal cancer. Skin cancer can be majorly [10] classified into two types Benign and Malignant. Benign is cancer that poses a lower level risk, whereas malignant is most harmful and affects all vital body organs. Cancer can be curable if it is detected at an early stage, via laser techniques or other medications. Whereas if detected at a later stage, it is difficult to cure as the cells can spread to other body parts.

II. RELATED WORK

The skin conditions expert group of the Global Burden of Disease 2010 study estimated the global burden of skin conditions worldwide they surveyed in 187 countries [1] and they found it all skin conditions combined were the fourth leading cause of non-fatal disease. Health can be lost due to 15 skin diseases [1]

Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine In this research paper, an intelligent diagnosis scheme is proposed for multi-class skin lesion classification. [5] The proposed scheme and is implemented via a hybrid approach. i.e. using deep convolution neural network and error-correcting output codes (ECOC) using support vector machine.

In this, the experiments were performed on 9,144 images which were taken from various sources. Alex NET, which is a pre-prepared CNN model and is utilized to separate the highlights. For the classification purpose, the ECOC SVM classifier was used [6].

Using ECOC SVM, the overall accuracy which was achieved is 86.21%. Their proposed intelligent expert system has been capable of classifying various skin types such as healthy, acne, eczema, benign and malignant skin lesions [7].

In another paper, [11] They collect the data by using a Mobile app and applied a convolution neural network to train and test the data. They proposed the model on PHDB melanoma dataset. They considered 60% fewer parameter compared to the base model so it more efficient by using on a mobile application. The proposed model was implemented on frameworks such as Keras and TensorFlow.

By using advanced computer based technology mobile teledermoscopy (MTD) helps to detect the patch on the skin is whether a cancerous or non-cancerous patch[11]
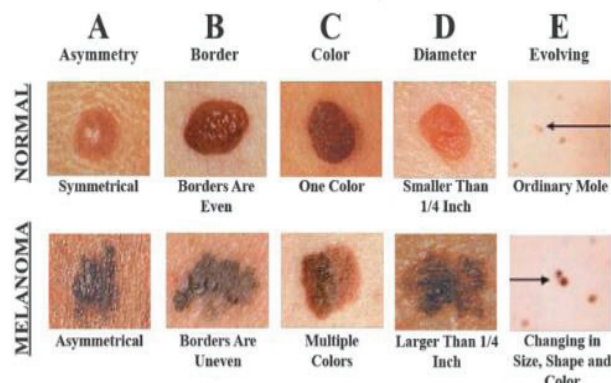

Fig. 1. Melanoma classification [11]

Dataset consists of images of malignant and benign lesions, where the images if both classes are in balanced amounts. i.e. PHDB dataset is divided into two equal parts ,one part included of benign images and the remaining are malignant images. It is partitioned as follows: 80% for training, 10% for validation, and 10% for testing.

Melanoma comprises as a most risky sort of disease fig 1. It affected more than 10,000 people in the US, so it's necessary to cure it as early as possible. It can help save more than 95% of people. Dermoscopy catches high resolution pictures of the skin getting away from the interference of surface reflections.

Past arrangement procedures are less fruitful for two significant reasons:

1) The past database isn't on a par with enough. Prior work depends on a significant number of measures of Dermoscopy skin sore pictures, so programs can't learn and separate helpful highlights.

2) The machine and technologies have limited computing ability, so it's difficult to deal with huge amounts of images.

Here they are attempting to build up a novel profound learning based answer to take care of melanoma characterization issues. In detail, I am attempting to take care of two significant issues:

1) The skin lesion recognition problem: U-Net 6 based neural network is used to achieve good detection and recognition accuracy.

2) The melanoma judgment: they use the exchange learning technique dependent on Google very much prepared Inceptions_v3 neural system.

Image Preprocessing: here they raised all the images into 150X150 for both skin_segmn and skin_recnn.

III. PROPOSED SKIN CANCER PREDICTION USING CNN

The proposed system will be implemented using deep convolutional neural networks (CNN) which will be capable of classifying skin lesions. We used a convolutional neural network (CNN) model of deep learning for the detection of cancer through skin cancer images. The proposed CNN model excelled will be implemented on frameworks such as Keras and TensorFlow. Dataset is divided into seven classes. HAM 10000 dataset is used for achieving better accuracy. Preprocessing is done to filter the data and thus avoid missing values, and resize larger images and crop if necessary. Training the model from scratch will require higher time and computing power, so Transfer Learning can be used, so for faster and better accuracy the MobileNet model can be used.Tto get a good accuracy a huge data set is
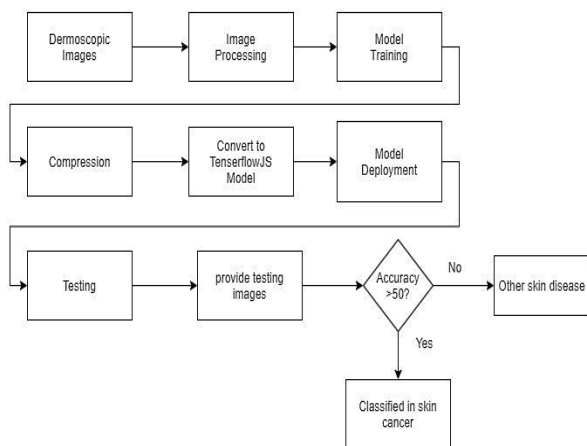

Fig. 2. Architecture of cancer detection web system

needed we do not have a huge dataset so prediction may not get that much with good accuracy. So using some additional datasets to train the data. Dataset of ImageNet trained over 14 million images [9]. We select a model that has been trained with a huge dataset. It helps to reduce training time and helps to improve accuracy.The Keras model formed is larger in size around 1 GB , The web app is made and the compressed model is thus connected .The UI design thus will enable the user to   so not suitable for web application so should be compressed, into a TensorflowJS model. The tensorflowJS library is used in python; this produces the shard files and json file. The web app is made and the compressed model is thus connected and the user can upload an image of the skin and will get the result in a few seconds.

A. Problem Statement

Skin diseases happen due to many causes. Skin diseases can create issues in the body parts. Many skin diseases may look the same at the initial stage. Skin cancer may start from a lump or spot and patchy skin psoriasis also looks like at the initial stage. Due to skin diseases, many problems may occur in life people may not live their daily life properly. Early detection of skin cancer and treating it well so many people can live their life stress-free. So to minimize this cause we make a CNN model to predict skin cancer in an early stage.

B. Proposal Steps:

1) Data Collection:

The training of neural networks for the skin lesion identification of cancer images is quite difficult due to the unavailability of the proper and good size of the dataset of the cancer images. As a solution for this HAM10000 ("Human Against Machine with 10000 training images") dataset is available on Kaggle which is a collection of dermatoscopic images from different populations, acquired and stored by different modalities. The dataset comprises of 10015 dermatoscopic images which is a good amount for training the machine learning models and can be used for academic purposes [13]. The dataset consists of a   vast collection of all the important diagnostic categories in the realm of pigmented lesions: Actinic keratosis and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc).

2) Image pre-processing:

Images of the dataset are segregated into 2 sets namely the training and the validation set, according to the 7 classes of skin cancer categories. The next step is to increase the amount of dataset thus for yielding better accuracy in our model the augmentation is done using the Image Data Generator function of Keras rotation, width shift, height shift, zoom, horizontal flip, vertical flip are done on the images so the no images in each of the 7 classes are approximate of the same amount. This is to be done as the model might encounter difficulty in training with the classes with fewer images and thereby reducing the accuracy achieved.

3) Training:

For developing deep learning models, data or dataset plays an important role. For deep learning models, a huge

2

dataset is the basic requirement as more data will give us the opportunity to predict results more accurately. As per our study, we required at least 50K images to train our model. [13]As in deep learning, tons of layers take lots of time to process the data and it will be easy if we execute it on GPU. We have applied transfer learning here as in transfer learning, we can use a pre-trained model and do the changes as per our dataset. So, instead of creating a new model, we can apply changes in the model such as we can alter layers or add layers or freeze a few layers which will give us more clarity on prediction of the result. For our project we used MobileNet architecture due to its high speed and low computing power, thus it is suitable to run on mobile platforms too.

The training was done on a TPU enabled Google Colab notebook as it is open source and also is   powerful and has great training speed. We need to choose how many layers we want to be trained. Here we are freezing the weights of all layers except the last 23 layers in the new model. The last 23 layers of the model will be prepared. We used Adam Optimizer. The learning rate was kept low 0.00001 and was ran for 30 epochs

Best Epoch Statistics:

```
val_loss: 0.5983754768871186
val_cat_acc: 0.8081023454157783
val_top_2_acc: 0.9125799573560768
val_top_3_acc: 0.9626865671641791
```

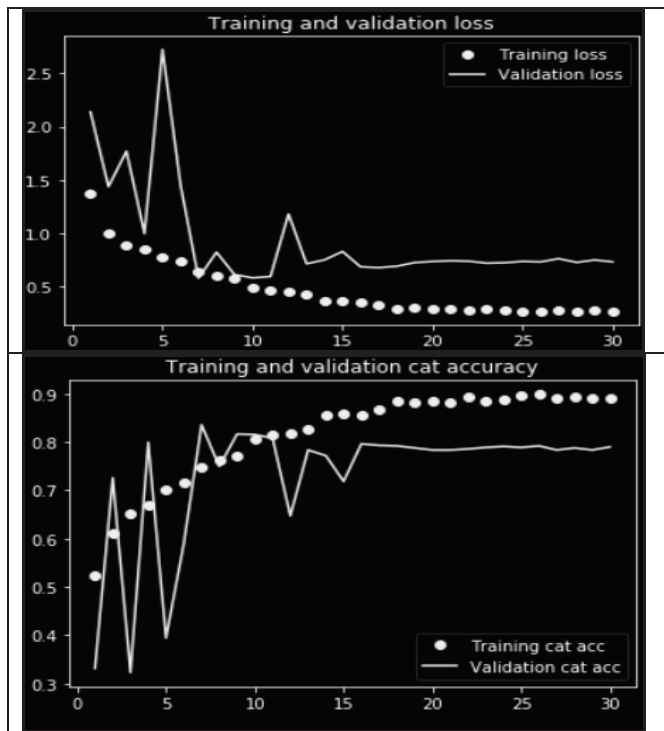Fig. 3.   Epoch statistics

Plot the Training Curves

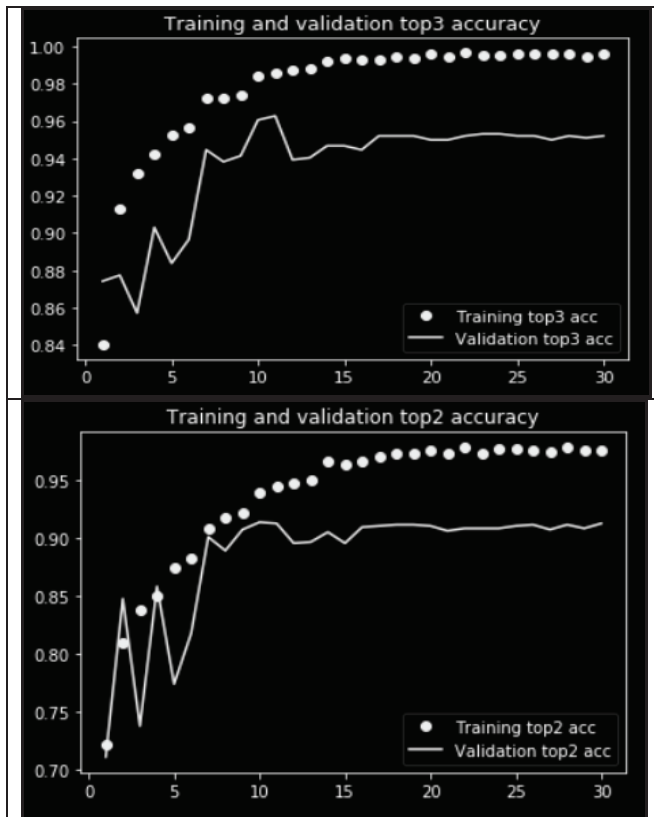Fig. 4.   Training and validation loss and accuracy

Fig. 5.   Training and validation loss and top2 and top3 accuracy

Accuracy obtained:

	precision	recall	f1-score	support
akiec	0.50	0.42	0.46	26
bcc	0.41	0.87	0.55	30
bkl	0.77	0.13	0.23	75
df	0.10	0.50	0.17	6
mel	0.28	0.49	0.35	39
nv	0.94	0.91	0.92	751
vasc	0.67	0.73	0.70	11
avg / total	0.86	0.81	0.81	938

Fig. 6.   Precision, recall and f1 score

4)   Compression and Conversion to tensorflowJS model:
The CNN model trained is very large in size, thus making it difficult to deploy it in on web, Thus we will compress the model to suit the web platform. Thus compression is done, using TensorflowJS library is used to convert the large model into compact size.

C.   Result
The model thus made will be able to predict the skin cancer in the image and give the probability of the classes which it can belong to , the classes namely are Actinic Keratosis, Basal Cell Carcinoma, Benign Keratosis, Dermatofibroma, Malignant Melanoma, Melanocytic Nevi[12], Vascular Lesions with an accuracy of around 85%. The user will be able to upload an image of the skin cancer

3

through the web interface and will get the results within a few seconds on the top prediction of the first 3 classes.

Output:

Cancer

Choose File | ISIC_0029292.jpg

Predictions

1. nv, Melanocytic Nevi: 0.996902

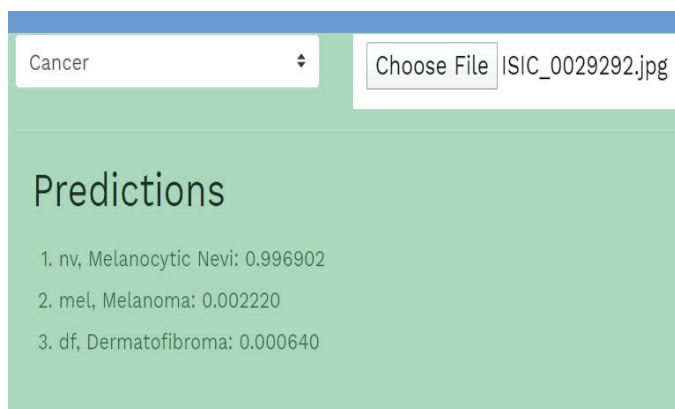2. mel, Melanoma: 0.002220

3. df, Dermatofibroma: 0.000640

Fig. 7. Screenshot of system

IV. CONCLUSION

Thus, a Skin cancer detection system will be implemented using deep learning methods and can be able to identify types of skin cancer tumor as benign or malignant. Future Scope: More and more images can be collected and trained to improve accuracy. New architecture can be implemented to minimize errors and reduce the training time and cost.

REFERENCES

[1] Hameed, Nazia & Shabut, Antesar & Hossain, Alamgir. (2018). Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine. 1-7. 10.1109/SKIMA.2018.8631525.

[2] Jaworek-korjakowska, J., & Kleczek, P. (2018). eSkin : Study on the Smartphone Application for Early Detection of Malignant Melanoma. Wireless Communications and Mobile Computing, 2018. https://doi.org/https://doi.org/10.1155/2018/5767360

[3] Salem, C., Azar, D., & Tokajian, S. (2018). An Image Processing and Genetic Algorithm-based Approach for the Detection of Melanoma in Patients. Methods of Information in Medicine, 57(1), 74–80. https://doi.org/10.3412/ME17-01-0061

[4] Picardi, A. (2013). Suicide risk in skin disorders. Clinics in Dermatology, 31, 47–56. https://doi.org/10.1016/j.clindermatol.2011.11.00

[5] N. Hameed, A. M. Shabut and M. A. Hossain, "Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine," 2018 12th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), 2018, pp. 1-7, doi: 10.1109/SKIMA.2018.8631525.

[6] N. Hameed, A. Ruskin, K. Abu Hassan and M. A. Hossain, "A comprehensive survey on image-based computer aided diagnosis systems for skin cancer," 2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), 2016, pp. 205-214, doi: 10.1109/SKIMA.2016.7916221.

[7] 2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA) (pp. 205–214). IEEE. https://doi.org/10.1109/SKIMA.2016.7916221

[8] Mathworks, Transfer learning-Matlab & Simulink, Retrieved August 7, 2018,from Omar Abuzaghleh, Buket D. Barkana, Miad Faezipour.

[9] Noninvasive Real-Time Automated Skin Lesion Analysis System for Melanoma Early Detection and Prevention", IEEE Journal of Translational Engineering in Health and Medicine, 2015

[10] Nazia Hameed, Antesar Shabut, Fozia Hameed, Silvia Cirstea, Alamgir Hossain. "Achievements of neural network in skin lesions classification", Elsevier BV, 2021

[11] Phillip Ly, Doina Bein, Abhishek Verma. "New Compact Deep Learning Model for Skin Cancer Recognition", 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2018

[12] Skin Diseases Classification Using Deep Leaning Methods, Curr Health Sci J. 2020 Apr-Jun; 46(2): 136–140. Published online 2020 Jun 30. doi: 10.12865/CHSJ.46.02.06

[13] Convolutional Neural Networks Using MobileNet for Skin Lesion Classification, July 2019 DOI:10.1109/JCSSE.2019.8864155,Conference: 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)