### Melanoma Prediction using Convolutional Neural Networks

Avantika Shenoy Department of Computer Engineering Thadomal Shahani Engineering College Mumbai, India Bhavin Shewkani Department of Computer Engineering Thadomal Shahani Engineering College Mumbai, India Tarun Utamchandani Department of Computer Engineering Thadomal Shahani Engineering College Mumbai, India

#### ABSTRACT

Melanoma is one of the most common types of skin cancer. Early detection of melanoma could prove to be the vital difference. The system proposes the use of neural networks to get a quick and accurate estimate and also determine the probability of skin cancer. The use of convolutional neural network as well as back propagation neural networks ensures that the system will have increased accuracy while the image processing techniques (RGB to grayscale, median filtering, segmentation) are in place to reduce the complexity and requirements of the system.

#### Keywords

CNN-Convolutional Neural Network, BPNN-Back Propagation Neural Network, SVM-Suport Vector Machines, Melanoma- a type of skin cancer that occurs in Melanocytes, Skin cancer-abnormal multiplication of skin cells, Melanocytes-skin cells that impart colour to the skin.

#### 1. INTRODUCTION

Year after year, ample of medical records are generated. These medical records can be of great use for biomedical informatics and machine learning purposes. Machine learning in itself can be used to detect and make predictions on the occurrences of many diseases. One such disease that can be taken into consideration is cancer. Cancer is the abnormal multiplication of the body cells and is life-threatening. There are various kinds of cancer depending on the parts of the body they are affecting. Skin cancer is the type of cancer that affects skin cells. There are 3 types of skin cells namely, squamous, basal and melanocytes. The type of cancer that affects melanocytes is called melanoma and is discussed in this paper. Melanoma is primarily detected through moles. In very rare cases melanoma affects the intestines. For women, melanoma moles occur most commonly on the legs and for men these moles occur most commonly on the back. Melanoma can be fatal but if detected early then it can be treated very effectively. Extensive research has been carried out on moles that could be malignant. For detection purposes, the asymmetry, border, colour, diameter and evolution of moles is taken into consideration. Moles that are malignant are asymmetric, have irregular borders, have different shades, are bigger than the diameter of an eraser at the back of pencils and they evolve. The proposed system discussed in this paper takes the irregularity of the mole's border as the key factor in determining whether the mole is cancerous or not. The system proposed in this paper makes use of convolutional neural networks followed by back propagation neural network..

### 2. REVIEW OF LITERATURE

'An Analysis of Skin Cancer Detection Using Imagery Techniques' [1]is a research paper by Chetna Garg et.al. The aim is to help aid the rapid increase in the number of melanoma skin cancer patients. They use data mining and Epiluminescence image processing. They also make use of the Adaptive Neuro-Fuzzy Inference System (ANFIS) Algorithm which provides better results than with other algorithms for diagnostic systems. The paper however, is more of an analysis of different research paper and suggests solutions, such as adaptive neuro-fuzzy inference system (ANFIS) algorithm, however, such techniques will be avoided as it is not offering a clear-cut solution.

Research paper by Shivangi Jain et.al. 'Computer aided Melanoma skin cancer detection using Image Processing'[2] tries to create a system to detect Melanoma using image processing and feature extraction. It starts off by taking the input image of skin lesion. This can be obtained in any lighting condition though any sort of camera could be used. The contrast of the image is then enhanced. Image Segmentation follows this. This image segmentation is performed using the automated thresholding proposed in the paper. Masking operations in R, G, and B planes are performed. Binary masks for each plane are obtained using thresholds which are then combined to produce a final lesion mask. 3 plane masking procedure was used to increase segmentation accuracy. The basis of this is to find the biggest blob in order to locate the skin lesion. Succeeding this is the application of edge detection. The main prerequisite for extracting the features is that the lesion must be separated from the surrounding normal skin. But the segmented image may contain other smaller blobs which are not the skin lesion. This is why, to avoid confusion, the biggest blob is taken into consideration. This provides the segmented image containing only skin lesion. The most prominent aspect of the Melanoma skin lesion is its Geometric Feature. Therefore, the extraction of this Geometric Feature takes place. Standard geometry features such as area, perimeter, diameter, circularity index, irregularity index is used. These features are used to get the classification and finally the result. However, the system proposed by them can be further improved by making use of neural networks to provide a far more accurate prediction. As it currently stands, it makes use of formulae that are very rigid and will not be able to detect certain shapes as cancerous. The neural network will be able to study the examples, and generate a better prediction.

Research paper by Abrham Debasu Mengistu et.al. is titled 'Computer Vision for Skin Cancer Diagnosis and Recognition using RBF and SOM.'[3] The aim of this was to detect the type of skin cancer. The image of the suspected area affected by skin cancer is taken. This image is then processed to get the enhanced image. so as to reduce noise. Analysis is performed on the image, and based on this analysis, features are extracted. Median filtering and feature extraction are used. Fifteen features are targeted for extraction which acted as the input features. Accordingly, a pattern was recognised. The classification is then done and the type of skin cancer is recognised. While they use support vector machine and median filtering, neural networks still give more accurate responses. Moreover, this only gives us the type of skin cancer and not the probability of skin cancer.

'Melanoma skin cancer detection and classification based on Supervised and Unsupervised Learning [4]' authored by Ms. H. R. Mhaske et.al, makes use of k-means clustering, support vector machines and back propagation natural network to detect whether or not a mole is cancerous. Various image processing techniques are used prior to feeding the images to get classified. The paper makes use of segmentation and median filtering to process the images. Feature extraction here focuses on the border of the moles. Median filtering is followed by segmentation. There are a lot of unwanted parts in an image that need to be processed. One such example of an unwanted part could be the hair. Median filtering makes sure that it removes all these features and focuses only on the parts that are necessary. Median filtering filters out the noise in an image. It thus improvises the image quality and helps vield better results during the latter stages of processing. The chances of attaining higher accuracy can be significantly improved by performing median filtering and thus retaining a smooth image. Post median filtering, segmentation is applied on the smooth image that has been obtained as a result of median filtering. Segmentation makes the analysis of the image easier by representing the images into something that is more meaningful. Segmentation focuses on the lines and curves in an image. In this paper and its proposed system, the focus is laid on the border of the image. Everything else is nullified. Thresholding technique is used to implement segmentation wherein the pixel values over the threshold values are set to 1 and those below it is set to 0. Melanoma features usually lay within the border of the mole. Classification algorithms such k-means clustering, support vector machines and back propagation neural network are then used to classify the moles as cancerous and noncancerous. K-means algorithm gave an accuracy of 52.63%, Support vector machine gave an accuracy of 80%-90% and back propagation neural network gave an accuracy of 60%-75%.

'The Melanoma Skin Cancer Detection and Classification using Support Vector Machine' is a research paper by Hiam Alguran et al. makes use of support vector machine (SVM) for classification. The steps involved in this study are segmentation using thresholding, feature extraction using gray level co-occurrence matrix, asymmetry, border, color, diameter (A, B, C, D) etc. The result achieved in this paper was 92.1%. SVM is based on the concept of decision planes that define decision boundaries. A decision plane separates the objects into distinguished classes. The simplest SVM plane is linear whenever the data can be linearly separated. But the data here is not linearly separable. therefore, kernel SVM with radial basis function was used to classify the data into benign or malignant. In the training stage, the classifier model was built using cross validation procedure to find the optimize parameters of the hyperplane to avoid biasing with overfitting.

'Pigment network-based skin cancer detection' authored by Naser Alfred et.al. uses Artificial Neural Network (ANN) for classification. The system proposed in this paper consists of 4 main stages viz. pre-processing, pigment network extraction, feature extraction and classification. The pre-processing stage of the proposed system enhances the original image by removing air bubbles, artefacts caused by applied gel before capturing images in addition to hairs, and other noise. The classification stage of the proposed system aims to reach a decision on the skin image type using a machine learning technique. In this paper, the extracted features have been used to feed an ANN as a binary classifier; i.e. the target will be 0 for abnormal and 1 for normal images respectively. The processes of classification were repeated six times according to the number of selected features and a cross-validation (CV) strategy has been used. A two-layer ANN (a single hidden layer and an output layer) has been used with 120 neurons in the hidden layer and the Radial Basis Function (RBF) as an activation function.

In conclusion to the literature review, three of the papers were selected and further analyzed to be the basis for the proposed system. The research paper by Shivangi Jain et. al. uses image processing[9] using geometric features which was promising in reducing the computational time. However, since neural networks was not used, the processing time and power of this system would be very high. The paper by H.R.Mhaske et. al. was also selected for further analysis. It compares and contrasts supervised and unsupervised learning methods for classification while systematically providing accuracy of each algorithm (i.e. K-Means, SVM, BPNN) thoroughly. However, there is a huge computational complexity for BPNN. To reduce this, CNN will also be implemented in the proposed system. Moreover, the output of this system was binary response and would say yes or no for the chances of melanoma. The proposed system will be an improvement of this with a percentage probability response. The paper by Md. Mahmudur Rahman et. al. was the final paper selected. There was fast and automatic low complexity segmentation. A high accuracy was also managed. However a dataset of just 360 images was used for this. Seeing these three papers, the image processing techniques will be mirrored and the limitations in the neural network aspects will be improved on.

#### **3. PROPOSED SYSTEM**



Fig 1: Proposed System

#### 3.1 Image

129450 images of resolution 480px are used in the system.80 percent of the images are used for training while the

remaining 20 percent is used for testing. The images are of moles which are both cancerous and non-cancerous.

#### 3.2 RGB to Gray

For image processing in many implementations, color information is not helpful in identifying important edges or other features. There are exceptions to this though. If there is an edge in hue that is hard to detect in a grayscale image, or if identification of objects of known hue is needed, then color information could be useful. If color is not needed then it can be considered as noise

#### 3.3 Median Filtering

Median filter[10] can be defined as a non-linear digital filtering technique, often used for removing noise from an image or a signal. This kind of noise reduction is used in the early processing stages and is helpful to the later stages i.e. segmentation for this system. Digital image processing used median filtering as it helps in removing edges and in noise reduction. In general, the median filter goes through the signal each entry at a time and compares them to the neighbouring entries and picking out the median. The pattern

of neighbours can be called as the "window", which slides, entry by entry, over the entire signal. For one dimensional signal, the few preceding and succeeding entries can be selected as the window. However, for two dimensional or multidimensional signals such as images, more complex window patterns are possible such as box or cross patterns. For a signal with an odd number of entries, finding the median can be fairly easy. The process is to simply sort the entries in ascending or descending order and select the central entry.

#### **3.4 Segmentation**

The partitioning of an image to obtain the area of interest in it is known as segmentation. Segmentation starts with a technique known as thresholding. In Thresholding, the peak value for skin and peak value for lesion is determined and threshold is selected in between these two peak points. Whichever pixels have intensity values which are greater than the threshold value is set as 0 whereas intensity values which are less than the threshold value are set as 1.

## 4. CONVOLUTIONAL NEURAL NETWORK

The proposed system will be using 5x5 kernels and 2x2 max pooling matrices. This kind of neural network[11] is used for the sparsity of connections that it uses. It thus lessens the complexity of the system. Edge detection is made easy with the use of convolutional neural network. After the preprocessing steps the image dataset is fed into the CNN. The rectified linear unit (RELU) activation function is used to make all negative pixel values zero. Thus, the matrix of the image pixels and the matrix of the kernel is convoluted. The result is then passed through the activation function. The result of this step then undergoes max pooling wherein the dimensions of the matrix are significantly reduced for quicker computation. The resultant matrix is then flattened and a softmax function is applied to the flattened values. Applying the softmax function gives probability values which can then be converted to a percentage in order to determine the chance of a person getting melanoma.





# 5. BACKPROPAGATION NEURAL NETWORK

Using the feedforward CNN will most likely give errors in the labelling of the images. In order to minimise the errors, a back propagation neural network is trained. In this kind of neural network, the weights are adjusted according to the errors and thus errors are largely minimised.



#### 6. CONCLUSION

Melanoma, if predicted beforehand can save quite a few lives. The paper suggests a system that makes use of convolutional neural networks to predict the percentage chance of a person getting affected by melanoma. The proposed system uses convolutional neural networks followed by back propagation neural networks that is expected to attain an accuracy of over 90%

#### 7. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the template.

#### 8. REFERENCES

- [1] Chetna Garg et.al.," An Analysis of Skin Cancer Detection Using Imagery Techniques" 2015, IJARCSSE.
- [2] Shivangi Jain et.al., "Computer aided Melanoma skin cancer detection using Image" International Conference on Intelligent Computing, Communication & Convergence (ICCC-2015)
- [3] A.D. Mengistu, D.M. Alemayehu. Computer Vision for Skin Cancer Diagnosis and Recognition using RBF and SOM, International Journal of Image Processing (IJIP), Vol. 9, Issue 6, pp. 311- 319, 2015.
- [4] Ms. H. R. Mhaske et.al., "Melanoma skin cancer detection and classification based on Supervised and Unsupervised Learning"
- [5] Ekta Singhal et.al., "Skin Cancer Detection using Artificial Neural Network" Volume 6, No. 1, Jan-Feb

2015 International Journal of Advanced Research in Computer Science

- [6] Md. Mahmudur et.al., "A Multiple Expert-Based Melanoma Recognition System for Dermoscopic Images of Pigmented Skin Lesions" Conference Paper · October 2008
- [7] Omar Abuzaghleh Buket et.al., "SKINcure: A Real Time Image Analysis System to Aid in the Malignant Melanoma Prevention and Early Detection" IEEE J Transl Health Med 2015
- [8] Jain S, jagtap V, Pise N. Computer aided melanoma skin Cancer detection using image processing. Procedia Computer Science. 2015;48:735–40.
- [9] Charu Khare , Kapil Kumar Nagwanshi, "Image Restoration Technique with Non Linear Filter", International Journal of Advanced Science and Technology, Vol. 39, February, 2012
- [10] James C. Church, Yixin Chen, and Stephen V. Rice Department of Computer and Information Science, University of Mississippi, "A Spatial Median Filter for Noise Removal in Digital Images", IEEE, page(s): 618-623, 2008.
- [11] N. Murata, S. Yoshizawa, and S. Amari, —Learning curves, model selection and complexity of neural networks, in Advances in Neural Information Processing Systems 5, S. Jose Hanson, J. D. Cowan, and C. Lee Giles, ed. San Mateo, CA: Morgan Kaufmann, 1993, pp. 607-614