Skin Cancer Classification Using CNN Classifier With Pytorch

Santhiya M[1], Mr. E. Ranjith[2]
[2] Assoc.Prof.
[1, 2] Krishnasamy College of Engineering and Technology, Cuddalore

Abstract- *In the past 10-year period, from 2010 to 2020,the annual number of melanoma cases has increased by 53%, partly due to increased UV exposure which leads to skin cancers. Melanoma is one of the most lethal types of skin cancer among other skin cancers, a fast diagnosis can lead to a very high chance of survival. The first step in the diagnosis of a malignant lesion by a dermatologist is visual examination of the suspicious skin area. A correct diagnosis is important because of the similarities of some lesion types; moreover, the diagnostic accuracy correlates strongly with the professional experience of the physician. Without additional technical support, dermatologists have a 65%-80% accuracy rate in melanoma diagnosis.*

The problem of classifying skin cancers has also moved into the focus of the machine learning techniques. Machine learning algorithm follows the classical workflow of machine learning techniques: preprocessing, segmentation, feature extraction, and classification.

I. INTRODUCTION

The skin is a vital organ that covers the entire outside of the body, forming a protective barrier against pathogens and injuries from the environment. But because it is located on the outer part, the skin is prone to disease. One of these diseases is known as skin cancer. Skin cancer is an abnormality in skin cells caused by mutations in cell DNA. One of the most dangerous types of skin cancer is melanoma cancer. Melanoma is a skin malignancy derived from melanocyte cells, the skin pigment cells that produce melanin. Because these cells are still able to form melanin, melanoma is mostly brown or black colored.

Common symptoms of melanoma are the appearance of new moles or changes in existing moles. Changes to the mole can occur due to exposure to ultraviolet light that damages the DNA of skin cells and genes that control cell growth and division resulting in the formation of malignant cells. One of the first steps to diagnosing melanoma is to do a physical examination using dermoscopy. With this dermoscopy examination, it can assess the size, color, and texture of moles that are suspected as melanoma. To determine a person with melanoma, a dermatologist conducts research from the results of dermoscopy examinations obtained and matches them with medical science to produce conclusions, but the detection weaknesses are strongly influenced by human subjectivity that makes it inconsistent in certain conditions.

II. LITERATURE SURVEY

i. Dermoscopy of pigmented skin lesions- a valuable tool for early diagnosisof melanoma(2001): This section paper is presented by Giuseppe Argenziano and H Peter Soyer and described this paper as dermoscopy opens up a new dimension in the clinical morphology of pigmented skin lesions and enables the well-trained Research with image-based can be maximized by utilizing information technology products, such as deep learning. Deep learning has become a hot topic discussed in the machine learning world because of its significant capability in modeling various complex data such as images and sound.

Convolutional Neural Network (CNN) is one of deep learning's methods that has the most significant result in image recognition because it tries to imitate the same way of recognizing images in visual cortex as humans so that they are able to process the same information.

The aim of this research is to build a system that can classify melanoma cancer through the images from the dermoscopy examination with Deep Learning training using the CNN method.

physician to improve the accuracy of diagnoses in general, and melanoma in particular. Digital follow-up examinations, teledermoscopy, and computer-aided or computer-assisted diagnosis of pigmented skin lesions are exciting new tools that will certainly change the management of pigmented skin tumours

ii. **Epidemiology of Skin Cancer Role of Some Environmental Factors (2010):** This Paper is presented by Gabriella Fabbrocini and described the paper as Climate changes, exposure to UVB and high levels of arsenic in drinking water, as well as several other environmental factors ,have been reported to be associated with melanoma, SCC and BCC. As the incidence rate of melanoma

iii. The Melanoma Skin Cancer Detection using Support Vector Machine (2017): This paper is presented by Hiam Alquran and described this paper as using the SVM based on the selected features from PCA, was successful in

III. PROPOSED APPROACH

The Proposed method does not require dermatologists to classify skin cancer types. It can be able to classify skin cancer by using the machine learning algorithm with better accuracy.

IV. DIFFERENT PLATFORMS

1. EXPLORATORY DATA ANALYSIS OF THE DATA SET: Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods. It helps determine how best to manipulate data sources to get the answers you need, making it easier for data and nonmelanoma skin cancer is dramatically increasing worldwide, a clearer understanding of causative factors is an essential step in their prevention. Therefore, further studies are required in order to investigate the potential effect of other possible risk factors and actuate prevention strategies based on avoiding them. classifying the extracted lesion ROI. The results of the SVM classifier show accuracy of 92.1%. These results explain the performance of the classifier model in discriminating the skin lesion into benign and malignant. scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions.

EDA is primarily used to see what data can reveal beyond the formal modeling or hypothesis testing task and provides a better understanding of data set variables and the relationships between them. It can also help determine if the statistical techniques you are considering for data analysis are appropriate. Originally developed by American mathematician John Tukey in the 1970s, EDA techniques continue to be a widely used method in the data discovery process today. The main purpose of EDA is to help look at data before making any assumptions. It can help identify obvious errors, as well as better understand patterns within the data, detect outliers or anomalous events, and find interesting relations among the variables.

2. TUNING RESNET 18 MODEL:

ResNet-18 is a convolutional neural network that is 18 layers deep. You can load a pre-trained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224.

a)Transfer learning is the reuse of a pre-trained model on a new problem. It's currently very popular in deep learning because it can train deep neural networks with comparatively little data. This is very useful in the data science field since most real-world problems typically do not have millions of labeled data points to train such complex models. Transfer learning has several benefits, but the main advantages are saving training time, better performance of neural networks (in most cases), and not needing a lot of data.

Usually, a lot of data is needed to train a neural network from scratch but access to that data isn't always available — this is where transfer learning comes in handy. With transfer learning a solid machine learning model can be built with comparatively little training data because the model is already pre-trained. This is especially valuable in natural language processing because mostly expert knowledge is required to create large labeled datasets. Additionally, training time is reduced because it can sometimes take days or even weeks to train a deep neural network from scratch on a complex task.

b)Oversampling and undersampling in data analysis are techniques used to adjust the class distribution of a data set (i.e. the ratio between the different classes/categories represented). These terms are used both in statistical sampling, survey design methodology and in machine learning.

c)Image data augmentation is a technique that can be used to artificially expand the size of a training dataset by creating modified versions of images in the dataset. Image data augmentation is used to expand the training dataset in order to improve the performance and ability of the model. Keras Image Data Generator helps you to augment your images in real-time while your model is still training! You can apply any random transformations on each training image as it is passed to the model. This will not only make your model robust but will also save up on the overhead memory.

d)In machine learning, it is a common practice to split your data into two different sets. These two sets are the training set and the testing set. As the name suggests, the training set is used for training the model and the testing set is used for testing the accuracy of the model.

While training a machine learning model we are trying to find a pattern that best represents all the data points with minimum error. While doing so, two common errors come up. These are overfitting and under fitting. The most common split ratio is 80:20.That is 80% of the dataset goes into the training set and 20% of the dataset goes into the testing set.

3.CNN MODEL TRAINING: CNNs are neural networks with a specific architecture that have been shown to be very powerful in areas such as image recognition and classification.

CNNs are a supervised learning method and are therefore trained using data labeled with the respective classes. Essentially, CNNs learn the relationship between the input objects and the class labels and comprise two components: the hidden layers in which the features are extracted and, at the end of the processing, the fully connected layers that are used for the actual classification task. Unlike regular neural networks, the hidden layers of a CNN have a specific architecture. In regular neural networks, each layer is formed by a set of neurons and one neuron of a layer is connected to each neuron of the preceding layer. The architecture of hidden layers in a CNN is slightly different. The neurons in a layer are not connected to all neurons of the preceding layer; rather, they are connected to only a small number of neurons. This restriction to local connections and additional pooling layers summarizing local neuron outputs into one value results in translation-invariant features. This results in a simpler training procedure and a lower model complexity.

i) Current Classifiers for Skin Lesions Based on Convolutional Neural Networks In this section, the individual CNN methods used to classify skin lesions are presented. CNNs can be used to classify skin lesions in two fundamentally different ways. On the one hand, a CNN pretrained on another large dataset, such as ImageNet can be applied as a feature extractor. In this case, classification is performed by another classifier, such as k-nearest neighbors, support vector machines, or artificial neural networks. On the other hand, a CNN can directly learn the relationship between the raw pixel data and the class labels through end-to-end learning. In contrast with the classical workflow typically applied in machine learning, feature extraction becomes an integral part of classification and is no longer considered as a separate, independent processing step. If the CNN is trained by end-to-end learning, the research can be additionally divided into two different approaches: learning the model from scratch or transfer learning.

A basic requirement for the successful training of deep CNN models is that sufficient training data labeled with the classes are available. Otherwise, there is a risk of overfitting the neural network and, as a consequence, an inadequate generalization property of the network for unknown input data. There is a very limited amount of data publicly available for the classification of skin lesions. However, through the use of a specific training procedure called transfer learning, powerful CNN models with several million free parameters can also be employed for classification, even if only a small amount of data are available for training. In this case, the CNN is pre trained using a very large dataset, such as ImageNet; it is then used as an initialization of the CNN for the respective task. In particular, the last fully connected layer of the pretrained CNN model is modified according to the number of training classes in the actual classification task. There are then two options for the weights of the pretrained CNN: to fine-tune all layers of the CNN or to freeze some of the front layers because of overfitting problems and to fine-tune only some back layers of the network. The idea behind this technique is that the front layers of a CNN contain more generic features (eg, edge or color-blob detectors) that are useful for many tasks, but the back layers of the CNN become increasingly specific to the details of the classes contained in the original dataset.

4.EVALUTING AND TESTING OUR MODULE:

This is how we can implement an end to end machine learning project based on computer vision. CNN is one of state of the art choice for image classification problems. We can utilize the transfer learning technique for quicker implementations and accurate results.

V. CONCLUSION

The aim of this project is to determine the accurate prediction of skin cancer. The CNN Model I created above was tested to recognize from an image whether a Skin Cancer cell is Benign or Malignant. The model has an accuracy rate on the training set of 96.7%, with a loss of .089 in the latest epoch, with that being the highest among all other epochs. The model has an accuracy rate on the test set of 71.51% with a loss of 1.4443 in the latest epoch. However, the model’s highest level of accuracy of 75.25% was in epoch 5/25.Overall, I would say the model has done well and has achieved my goal of being more than 70% accurate.

VI. FUTURE WORK

The Proposed method does not require dermatologists to classify skin cancer types.It can be able to classify skin cancer by using the machine learning algorithm with better accuracy.

REFERENCES

[1] J Abdul Jaleel, Sibi Salim, Aswin.R.B, Computer Aided Detection 01 Skin Cancer, International Conference on Circuits, Power and Computing Technologies, 2013

[2] Mariam A.Sheha ,Automatic Detection of Melanoma Skin Cancer, International Journal of Computer Applications, 2012.

[3] Santosh Achakanalli & G. Sadashivappa , Statistical Analysis Of Skin Cancer Image A Case Study , International Journal of Electronics and Communication Engineering (IJECE), Vol. 3, Issue 3, May 2014.

[4] Abrham debasu mengistu, dagnachew melesew alemayehu "computer vision for skin cancer diagnosis and recognition using rbf and som " international journal of image processing (ijip), volume (9) : issue (6) 2015.

[5] M.Chaitanya Krishna, S.Ranganayak ulu, Skin Cancer Detection and Feature Extraction through Clustering Technique, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 3, March 2016.

[6] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," vol. 542, no. 7639, p. 115, 2017.

[7] P. Tschandl, C. Rosendahl, and H. J. a. p. a. Kittler, "The HAM10000 Dataset: A Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions," 2018.

[8]]yuexiang liid and linlin shen, "skin lesion analysis toward melanoma detection using deep learning network" sensors mdpi 11 february 2018.

[9] Z. Ouhib *et al.*, "Aspects of dosimetry and clinical practice of skin brachytherapy: American Brachytherapy Society working group report," vol. 14, no. 6, pp. 840-858, 2015. [8] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, M. J. M. T. Lee, and Applications,"The skin cancer classification using deep convolutional neural networks," pp. 1-16, 2018.

[10] MohdAnas, Ram Kailash Gupta, Dr. Shafeeq Ahmad, "Skin Cancer Classification Using K Means Clustering", International Journal of Technical Research and Applications, Volume 5, Issue 1, 2017.

[11] Hutokshi Sui, ManishaSamala, Divya Gupta, Neha Kudu, ‘‘Texture feature extraction for classification Of Melanoma", International Research Journal of Engineering and Technology (IRJET), Volume 05, Issue 03, March, 2018.

[12] Vedanti Chintawar, Jignyasa Sangha vi, ‘‘Improving Feature Selection Capabilities in Skin Disease Detection System", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 8, Issue 8S3, June, 2019.