Classification of Skin Lesion for Cancer Detection

Dhairya Arora and Chhavi Bhadana

Indira Gandhi Delhi Technical University for Women

May 24, 2023

Abstract

The medical industry faces a significant challenge in recognising and observing skin diseases. The number of people with skin problems is growing at a faster rate as a result of increased pollution and the consumption of poor nourishment. Unfortunate skin not only harms the well-being of an individual but also hurts one's certainty. Skin checking is an important step towards early detection of any harmful or initial changes in the skin that may lead to a skin disease. If a patient with a pigmented lesion is diagnosed as having or at risk of getting skin cancer, immediate steps can be taken to reduce its risk or eliminate the malignancy if detected early. In order to detect cancer, a reliable and automatic skin lesion classification system is required. Machine learning technologies can help in increasing the capability of frameworks that can classify different types of skin diseases. To identify skin maladies, it is required to separate the benign and malignant lesions. In this paper, Convolutional Neural Networks (CNN) algorithm has been chosen and executed on pigmented skin lesions in dermoscopic images to anticipate the exact class of lesion as mentioned above. It was discovered that the chosen algorithm gave a high testing accuracy of 94.167%.

Keywords
Machine Learning, CNN, Skin Cancer, Dermoscopy

1 Introduction

Skin is the largest organ of the human body. It senses the environment to protect the internal organs and tissues from harmful microscopic organisms, pollution, and sun exposure. External and internal variables have an impact on the skin. Skin problems are influenced by an array of cause: internal factors that influence skin conditions are genetics, hormones and specific conditions such as diabetes; external factors include stress, diet, climate, sun exposure and skin products.
Occasionally one would self-diagnose and attempt to treat their skin problems with home remedies before consulting a professional. If those strategies aren't appropriate for that type of skin disease, they could have negative consequences. In view of the fact that skin diseases (bacterial, viral or fungal) have a tendency to be transmissible, they must be managed early on. For a professional diagnosis the doctor's experiences and subjective judgments are used to draw conclusions about the patient's symptoms. Human health may be jeopardised if the decision is incorrect or delayed. As a result, it is necessary to develop an effective approach for early detection and diagnoses of skin lesions.

In this project we came across one such health issue that affects many people around the world, SKIN CANCER. Skin cancer is an abnormal development of skin cells that occurs most commonly due to sun-exposure. Unrepaired DNA damage promotes mutations, resulting in an out-of-control development of aberrant cells in the epidermis. These mutations lead skin cells to grow rapidly, resulting in malignant tumours. Basal cell carcinoma, squamous cell carcinoma, and melanoma (the deadliest type of skin cancer) are the three most common kinds of skin cancers.

Analysis of lesions is an major part of medicine. Through medical research, clinic appointments, and hospital visits, the healthcare system acquires a massive amount of data from patients. This information can be studied to uncover and realize hidden facts and figures about a disease that could lead to a breakthrough in patient care, therapy, or diagnosis. Accurate data analysis can aid in the early identification of skin cancer, which may reduce the number of patients dying.

Image and pattern-based discovery of many skin diseases is possible with a variety of improvements. Machine learning is one of the fields that has the potential to play a significant role in the operative and precise detection of many types of skin diseases. Diseases can be classified using image classification algorithms.

Automated analysis using machine learning would create a system and framework in the medical field that would aid in improving clinical reliability, assisting physicians in communicating objectively, reducing human fatigue-related errors, lowering mortality rates, lowering medical costs, and identifying diseases more quickly. A machine learning method that can categorize both malignant and benign pigmented skin lesions is a step toward achieving these goals.

Image classification is a supervised learning problem in which a model is trained to recognise a large number of objective classes. Many machine learning and deep learning algorithms exist that can recognise and predict various skin disease types based on their classifications. Convolutional Neural Networks (CNN) is one such algorithm utilized in this research to classify various skin lesions into benign and malignant categories.

We compiled our own dataset by extracting various skin lesions' images from a few different datasets available online[1][2][3][4][5][6]. Using separate train and test datasets, further classified into benign and malignant. The training dataset contained a total of 1851 benign images and 1990 malignant images, which made it an almost balanced class distribution. The testing dataset contained 180 benign as well as malignant images. Fig 1 shows some of the sample images from our dataset.

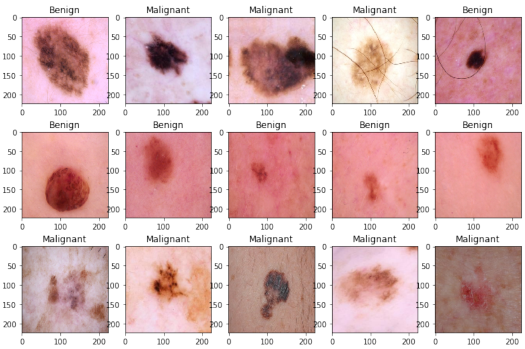

Figure 1: Images from our dataset

2 Literature Review

[7]Ilias Maglogiannis stated the importance of selecting a powerful variable for classification by acquisition techniques and feature extraction. However more systematic trials are needed because even a system with best classification had low performance in terms of accuracy and confidence in diagnosis.[8]Titus Josef Brinker1,2, MD, Achim Hekler1 mentioned that CNN algorithm displays a high performance with a state of art skin lesion classifier. The model learnt between input object and class labels and comprised of two components: a hidden layer, in which features were extracted and a fully connected layer that was used for actual classification.[9]Thurnhofer-Hemsi, K., Domínguez, E. Used two classification methods with CNN based on transfer learning. Five convolutional neural networks were applied on the HAM10000 dataset, and used in a hierarchical model with 2 levels - the first level dealt with distinguishing between nevi and non-nevi images and the second one to classified the malignant moles; and a plain model.[10]Tareq Tayeh et al presented a self-built model TSM12 and achieved an accuracy of 88.5% with VGG16 model. They used a total of 4 CNN models and kept the hyper parameters at default values. The dataset used, HAM1000, was imbalanced.[11]Shuchi Bhadula et al. aimed at adding machine learning methods to the improvement of capable frameworks which could order various classes of skin illnesses. Five diverse machine learning algorithms were applied on skin infection data set to detect the exact class of skin disease. The algorithms used were, Random forest, CNN, kernel SVM, naive bayes and logistic regression. Then an examination dependent on confusion matrix parameters and training accuracy was performed and delineated using graphs. It was discovered that CNN gave the best results for the right expectation of skin diseases among all selected.[12]M. Shamsul Arifin proposed a system for the dissection of skin diseases using color images without the need for doctor intervention. The system consisted of two stages, the first for detection of the infected skin by using color image processing techniques, k-means clustering and color gradient techniques for identification of the diseased skin and the second for classifying the disease type using artifi-

cial neural networks. The system was tested on six types of skin diseases with average accuracy of 95.99% for the first stage and 94.016% for the second.[13]Ahmed A. Elngar et al. aimed at creating a system combining Convolutional Neural Networks with a Support Vector Machine classifier to develop a mobile android application to detect skin diseases. The results obtained helped in detecting the skin disease and providing the user with a treatment-related prescription with high accuracy.[14] Adre Esteva et al compared melanoma vs benign predictions of their CNN model and that of actual dermatologists' on the same validation set. On average the CNN model performed better than the dermatologists'.

3 Background

3.1 CNN

CNN (Convolutional Neural Network) is a neural network with a unique design that has proven to be extremely effective in areas like image recognition and classification. CNNs have proved to be instrumental in recognizing faces, objects, and traffic signs accurately, and hence can be found in robots and self-driving automobiles.

neuron is connected to the neuron of the layer before it. In a CNN, the design of hidden layers differs slightly. A layer's neurons are not connected to all of the neurons in the preceding layer; rather, they are connected to a restricted number of neurons. This restriction to limited connections and additional pooling layers that sum local neuron outputs into a single value, results in translation-invariant features, which in turn results in a simpler training procedure and a lower model complexity.

CNN comprises of several kinds of layers:

• *Convolutional Layer*: This is the layer where filters are applied to the images.

• *Relu*: It is a correction layer which replaces all the negative values, received as inputs, with zeros. Hence acting as an activation function.

• *Pooling Layer*: It is used for down sampling. It cuts down on computing expenses and to some extent reduces over-fitting.

• *Dropout*: It is used to randomly drop some nodes while training. This helps the network to learn features in a distributed way. Thus, preventing over-fitting and improves generalization.

• *Flatten*: It is used to convert the data into 1-D array to input it into the next layer.

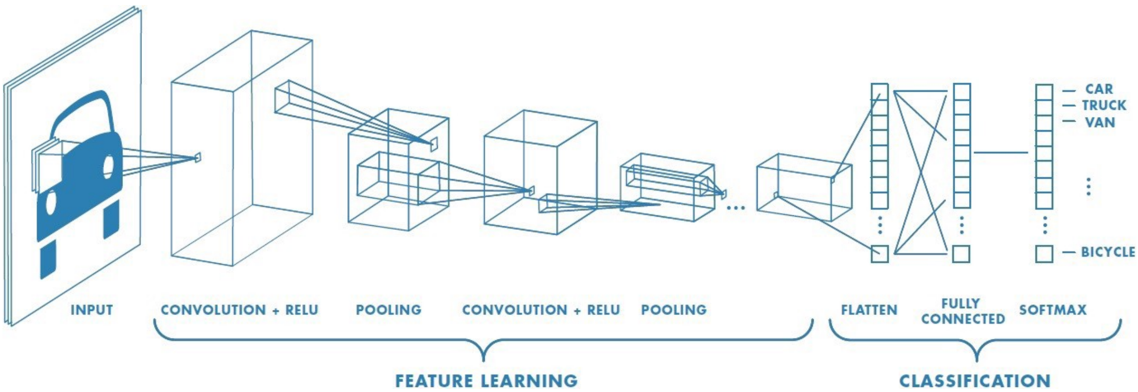

Figure 2: CNN

CNNs are supervised learning algorithms that are trained using labelled data with the appropriate classes. In essence, CNNs learn the link between the input objects and the class labels, and are made up of two parts: hidden layers in which the features are extracted, and fully connected layers at the conclusion of the processing that are employed for the actual classification task. The hidden layers of a CNN, unlike conventional neural networks, have a unique architecture. Each layer of a normal neural network is made up of a group of neurons, and each layer's

• *Fully Connected*: It takes the input from the final pooling or convolutional layer after it is flattened. This is a global layer, i.e., it can introduce any kind of dependence.

• *Softmax*: It is an activation function in the output layer of neural network model that is used to predict a multinomial probability distribution.

• *Dense*: A dense layer in a neural network is one that is deeply connected to the layer before it, meaning that the layer's neurons are connected to every neuron in the layer before it.

In artificial neural network networks, this layer is the most commonly used layer.

3.2 Optimizer

Optimizers are algorithms or methods for altering the characteristics of a neural network, such as weights and learning rate, in order to minimize losses and produce the best accurate results. Optimizers solve optimization problems by minimizing the function. The optimizer used in this research was Adam, which is one of the best optimizers for deep learning models. It has proven to be effective in training the neural networks in less time and more efficiently. It's a deep learning model training optimization approach that replaces stochastic gradient descent. Adam combines the best features of the AdaGrad and RMSProp methods to create an optimization technique for noisy issues with sparse gradients.

3.3 Workflow

The dataset was prepossessed for convenience in the code. The images were loaded in four separate directories and converted into numpy arrays. All images were normalized to be of the same size and the CNN model was built using the above discussed layers and Adam optimizer. Finally the built model was fit on the training dataset and tested on the validation dataset. Figure 3 shows a flowchart of the proposed methodology.

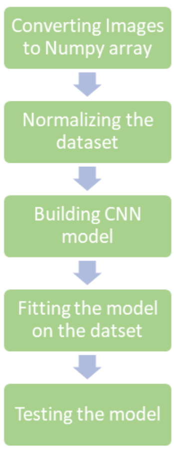

Figure 3: Flowchart

4 Results

The serious implications of a small lesion on skin can be understood from the above discussion. Convolutional Neural Networks have a marvellous architecture that helped us predict melanoma with an accuracy of 94.167 % on the testing dataset.

Figure 4 shows a sample prediction output of the model.

Figure 5 shows the model's progress overtime (epochs) on both the training and validation sets. In terms of model accuracy both steeped between 0-10 epochs and showed a consistent accuracy afterwards.

Figure 6 depicts the model loss on training and testing sets. While the loss increased during training, it decreased during testing.

Figure 4: Predictions sample

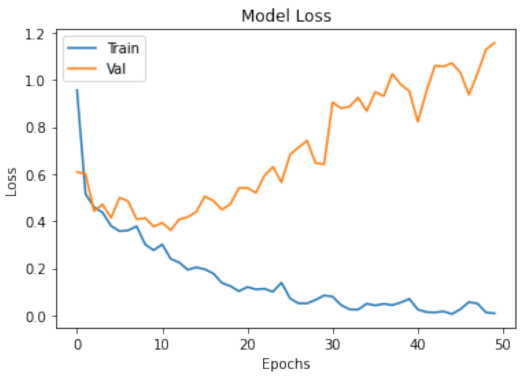

Figure 5: Epochs vs Model loss

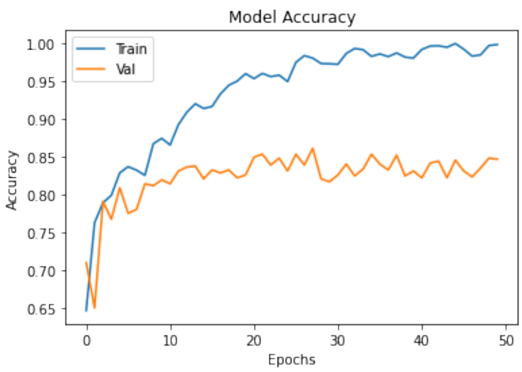

Figure 6: Epochs vs Model Accuracy

5 Discussion

An issue that occurs with the comparison of skin lesion classification methods is that the considered problem formulations of different works differ, sometimes only slightly. The problem occurs not only for the considered training classes and the data that is used, but also for the presented statistical quantities. Also, some works use private data of skin clinics in comparison to the publicly accessible data as used in this research.To create diversification we gathered images from datasets of different studies and regions; mostly light skinned. A similar model[15] that used a publicly available dataset achieved an accuracy of 81.36% whereas the dataset we created improved the accuracy to 94.167%. This displays the need for a diverse and balanced data to carry out similar studies.

6 Conclusion and Future Scope

Skin disease detection is a major problem in the medical industry. Detecting a skin lesion as benign or malignant at an early stage could help in reducing the chances or transmission and in some cases even prevent deaths. Machine learning algorithms have the potential to be useful in the detection of various skin diseases. In this research, with the help of convolution neural networks, we classified different skin lesions into benign and malignant. This would prove to be helpful for physicians and patients to accordingly act upon the necessary precautions and procedures for the treatment of the lesion. In the future, the system developed in this project can be enhanced to further classify the lesions to determine the most probable disease that a person has.

References

[1] Kaggle dataset : skin-cancer-detection-using-cnn.

[2] Kaggle dataset : datameln.

[3] Kaggle dataset : skin-cancer-malignant-vs-benign.

[4] Kaggle dataset : scdataset.

[5] Kaggle dataset : skincancerdataset.

[6] Kaggle dataset : melanoma.

[7] Ilias Maglogiannis and Charalampos N. Doukas. Overview of advanced computer vision systems for skin lesions characterization. *IEEE Transactions on Information Technology in Biomedicine*, 13(5):721–733, 2009.

[8] Titus Brinker, Achim Hekler, Jochen Utikal, Carola Berking, Alexander Enk, and Christof Kalle. Skin cancer classification using convolutional neural networks: Systematic review (preprint). *Journal of Medical Internet Research*, 20, 08 2018.

[9] Enrique Domínguez Karl Thurnhofer-Hemsi. A convolutional neural network framework for accurate skin cancer detection. *Neural Processing Letters volume*, 2021.

[10] Tareq Tayeh et al. Skin lesion classification using convolutional neural networks. 2020.

[11] Chitransh Kulshrestha, Piyush Juyal, Sachin Sharma, and Shuchi Bhadula. Machine learning algorithms based skin disease detection. 9:4044–4049, 05 2020.

[12] M. Arifin, Muhammad Golam Kibria, Adnan Firoze, Md Ashraful Amin, and Hong Yan. Dermatological disease diagnosis using color-skin images. *Proceedings - International Conference on Machine Learning and Cybernetics*, 5:1675–1680, 07 2012.

[13] Ahmed A. Elngar, Rishabh Kumar, Amber Hayat, and Prathamesh Churi. Intelligent system for skin disease prediction using machine learning. *Journal of Physics: Conference Series*, 1998(1):012037, aug 2021.

[14] Esteva et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542:115–, January 2017.

[15] Prakrati. Skin cancer detection using cnn.

[16] Chitransh Kulshrestha, Piyush Juyal, Sachin Sharma, and Shuchi Juyal. Machine learning algorithms based skin disease detection. 9:4044–4049, 05 2020.

[17] Nawal Soliman ALKolifi ALEnezi. A method of skin disease detection using image processing and machine learning. *Procedia Computer Science*, 163:85–92, 2019. 16th Learning and Technology Conference 2019Artificial Intelligence and Machine Learning: Embedding the Intelligence.

[18] R. Sumithra, Mahamad Suhil, and D.S. Guru. Segmentation and classification of skin lesions for disease diagnosis. *Procedia Computer Science*, 45:76–85, 2015. International Conference on Advanced Computing Technologies and Applications (ICACTA).

[19] Rahat Yasir, Md. Ashiqur Rahman, and Nova Ahmed. Dermatological disease detection using image processing and artificial neural network. In *8th International Conference on Electrical and Computer Engineering*, 2014.

[20] Avantika Shenoy, Bhavin Shewkani, and Tarun Utamchandani. Melanoma prediction using convolutional neural networks. *International Journal of Computer Applications*, 182:17–20, 12 2018.

[21] 3blue1brown. Neural networks, 2018.

[22] Unknown. Convolutional neural network (cnn), 2021.

[23] Mabrouk S.M Sheha A.M and Sharawy. Automatic detection of melanoma skin cancer using texture analysis. *International Journal of Computer Applications*, 42:22–26, 3 2012.

[24] Dr. Vidya Kharkar Ms. Seema Kolkur, Dr. D.R. Kalbande. Machine learning approaches to multi-class human skin disease detection. *International Journal of Computational Intelligence Research*, 14:29–40, 2018.

[25] Mukti E. Jadhav. Classification of dermoscopy images for early detection of skin cancer – a review. *International Journal of Computer Applications*, 178:37–43, 2019.

[26] Mrunal Bhosale, A Karle, Rutuja Sonawane, and Rutuja Wagh. Survey on machine learning approach for the detection of melanoma using dermoscopic images. 2018.

[27] Sonali N. Jadhav and Durgaprasad K. Kamat. Analysis and detection of skin cancer. *IOSR Journal of Electronics and Communication Engineering*, 9:50–54, 2014.

[28] Eijaz Allibhai. Building a convolutional neural network (cnn) in keras, 2018.

[29] Aromal Jose Baby. Building a cnn model with 95

[30] V.Priyadarshini J.Rubika A.T.Madhavi, S.Meeraa. Melanoma skin cancer detection using convolutional neural network. *ISSN*, 7, 20120.

[31] Aswin.R.B Dr. J. Abdul Jaleel, Sibi Salim. Diagnosis and detection of skin cancer using artificial intelligence. *International Journal of Engineering and Innovative Technology*, 3:311–315, 8 2013.