Sangyoon Kim

CPSC393: Machine Learning

Dr. Parlett

April 9, 2023

Introduction

Skin cancer is the most common form of cancer in the United States and accounts for more than 5 million cases each year. In addition, it is the most common type of cancer globally, and early detection and diagnosis are crucial for effective treatment. Developing an accurate and automated method for skin cancer detection using machine learning has great potential for improving patient outcomes. In recent years, deep learning models have shown ability in classifying skin lesions and aiding in the diagnosis of skin cancer. In this report, a deep convolutional neural network (CNN) model for binary classification of skin as cancer or non-cancer.

Analysis

The skin cancer binary classification dataset from Kaggle, which contains 288 images of skin lesions with corresponding labels indicating whether it is the cancer or the non-cancer. The dataset has a class imbalance, with 84 cancer samples and 204 non-cancer samples respectively. First, explored the distribution of labels in the dataset and found the class imbalance. Also, visualized some sample images from the dataset to get an idea of the types of skin lesions present. Then resized them to 224x224 pixels to standardize them for the model. In addition, observed some sample images from both classes and noticed that cancer image tends to be more irregular in shape and have more varied coloration compared to non-cancer image.

Methods

A deep convolutional neural network (CNN) built using the Keras with TensorFlow as the backend for binary classification of skin cancer. The model includes of 5 convolutional layers with the ReLU activation, followed by 3 max pooling layers. The activation function used in all layers except the output layer was the ReLU, and a sigmoid activation function used for the output layer to produce a binary classification output because get a probability value between 0 and 1. In addition, dropout added regularization applied to prevent overfitting the model, with a rate of 0.5 for all layers.

The model used data augmentation techniques to increase the size of the training data. The model used the binary cross-entropy loss function and the Adam optimizer to train the model. Also, the model trained for 100 epochs with a batch size of 32.

Results

The model had an accuracy of 88.8% on the training data for cancer images and 89.5% for non-cancer images. On the test data, the model had an accuracy of 86.3% for cancer images and 85.8% for non-cancer images. The confusion matrix and classification report suggested that the model had a slightly higher false negative rate for cancer images compared to non-cancer images, indicating that the model is better at correctly identifying non-cancer images than cancer images. The model performed almost well, achieving an accuracy of around 86% on both the cancer and non-cancer classes. It suggests that this model has potential as a screening tool for skin cancer, but further evaluation and optimization are needed to improve its performance.

Reflection

In this assignment, a deep CNN model built for binary classification of cancer or non-cancer. Through build a model, it showed the importance of data preprocessing and augmentation in improving the performance of a deep learning model. Also, how to build and optimize a deep convolutional neural network for image classification tasks. A limitation of the model was that it was trained and tested on a relatively small dataset, which may limit its generalizability. In the future, using transfer learning techniques would improve the performance of the model. Also, handle the class imbalance in the dataset and try different architectures and hyperparameters to further improve the model's performance, or test it on a larger dataset. Overall, this assignment has provided insights into the challenges and opportunities in this field and demonstrated the potential of deep learning models in aiding in the diagnosis of skin cancer.