A deep skin cancer classification approach using image and structured information

Supamdeep Singh
Department of Physics and Computer Science
Wilfrid Laurier University
Waterloo, Canada
supambains007@gmail.com

Jiashu (Jessie) Zhao
Department of Physics and Computer Science
Wilfrid Laurier University
Waterloo, Canada
jzhao@wlu.ca

***Abstract*— In this paper we describe an approach to implement a hybrid skin cancer detection machine learning model that uses the combination of two models. The final concatenated model makes prediction based on the information it gains from the two models. The first model being the Convolutional Neural Network model that learns from the image's dataset and the second Artificial Neural Network model that learns from the textual data associated with the images. Hence the final concatenated model uses these models to make predictions.**

Keywords*— *Convolutional Neural, Artificial Neural Network, Rectified Linear Unit, EfficientNetB3, LabelEncoder

I. Introduction & Context

Skin cancer is one of the deadliest forms of cancer and it is a major public health concern. According to the skin cancer foundation, more than 2 people die of skin cancer every hour [1]. Its detection plays a major role in its ability to impact patients' life. With any other form of cancer the sooner it is detected the better it is for the health of the patient. It is the most commonly diagnosed cancer in the United States [2]. The use of neural networks and machine learning can help medical professionals in diagnosing patients with skin cancer more accurately and quickly. There have been several studies done on the use of AI and machine learning in the diagnosis of skin cancer such as AI-powered computer-aided diagnosis which improves the accuracy of the medical exams of the affected skin region [3]. This report proposes a hybrid model which uses a combination of CNN and ANN which leverages the image data and tabular data to predict the type of skin cancer. There are six types of skin cancer that can affect human life which are (a) Basal Cell Carcinoma (BCC), (b) Squamous Cell Carcinoma (SCC), (c) Melanoma (MEL), (d) Actinic Keratosis (ACK), (e) Nevus (NEV) and (f) Seborrheic Keratosis (SEK).

II. Related Work

The majority of previous works in this field involve the classification of skin cancer, based on only the image dataset. The use of the CNN model helped achieved an accuracy of 89.5% solely based on the image data [4]. Other works include fine-tuning of pre-trained models such as Mobile-Net for classifying the skin cancer type with an accuracy of 83.1% [5]. With so many advancements in using image data to predict the type of skin cancer, there have been very few studies that benefit from the combination of tabular data with image data. One of the studies showed the use of multiple pre-trained models such as GoogleNet, VGGNet-13/19-bn, and ResNet with the EfficientNetB3 which is a deep learning model for extreme gradient boosting (XGB) to achieve an accuracy of 78% using the images and clinical data. The proposed model in this report uses a simplistic sequential CNN model concatenated with another sequential ANN to achieve an accuracy of 68% that leverages both the image and text/tabular data.

III. Methodology

The new proposed model uses sequential CNN and ANN to predict target features. Overall, the final predictive model consists of two segments. The first segment is the convolutional neural network which takes in images of input shape 128x128x3. The second segment is the basic sequential neural network which takes in 60 features of the metadata associated with the images. Finally, the model is concatenated to predict an outcome based on the image and the associated metadata.

The flow starts from the concatenated model where the final model takes in an array of image and associated 60 descriptive features. After taking in the image and the corresponding associated metadata, the model passes image data through the convolutional model consisting of convolutional layers and the metadata through the artificial neural network consisting of multiple dense layers. The information flows parallelly among both the sub models (CNN and ANN). Soon as the information is reached to the output layers of the both the models, then the information is concatenated. This concatenated model then passes the extracted information from both image dataset and the textual dataset through a series of dense layers and finally outputs the predicted value.

A. CNN Model Architecture

This section will cover the complete architecture of a convolutional neural network. The model takes in a labeled images from the dataset with a resolution of 128x128x3 with the activation function being Rectified Linear Unit (ReLU). The model has 14 layers which consist of 5 convolutional 2D layers, 5 Maxpool layers, 3 dense layers, and 1 flatten layer. We enabled zero padding by mentioning the padding to be the "same". The kernel size used throughout the layers was 3x3 with the loss function being categorical cross-entropy. The model has an output shape of (None,32). The architecture of CNN model can be seen in Fig. 3.1.

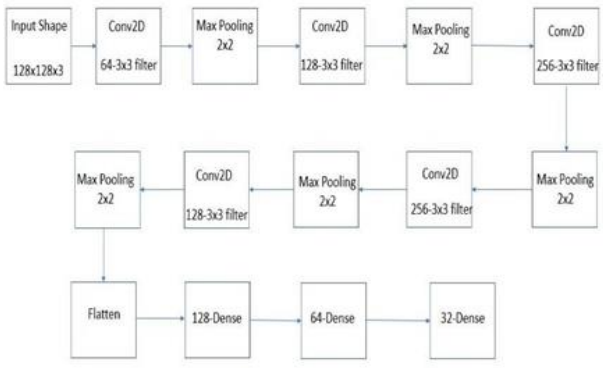


Fig. 3.1. CNN Model Architecture

The model was trained over 1470 labeled images with the number of epochs being 20. The images were labelled from the imageDataGenrator from the keras module. The model predicted the target features with an accuracy of 59% on the testing dataset and validation accuracy of 64% solely based on image data

B. ANN MODEL ARCHITECTURE

This section covers the details involving the artificial neural network which takes in 60 descriptive features from tabular/text data. This model consists of 4 dense layers with an activation function being Rectified Linear Unit (ReLU). The model uses a loss function of sparse categorical crossentropy. The model has an output shape of (None,256).

The model was trained over 1470 rows of data with the number of epochs being 100. The model predicted the target features with an accuracy of 71% on the testing dataset and a validation accuracy of 79%. The architecture of ANN model can be seen below Fig 3.2.

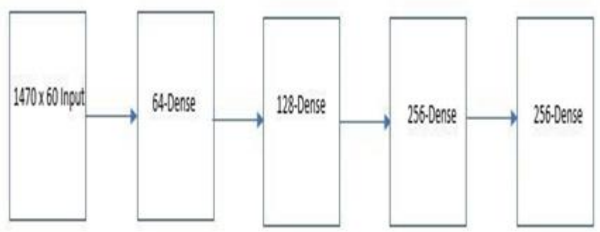


Fig. 3.2. ANN Model Architecture

C. CONCATENATED MODEL ARCHITECTURE

The final concatenated model is a concatenation of the first two models. This way the model leverages the information retrieval from both the textual and image data. The outputs from the CNN model from image data and ANN model from metadata converges to concatenated layer followed by three dense layers. This way the final concatenated model predicts the target feature with the rich information gained via both the models. The input shape of the final model is an array that takes in an image of resolution 128x128x3, and 60 descriptive features associated with it. The final model has a training accuracy of 85% and the testing accuracy of 68%. The architecture of this model is shown at the bottom of the paper.

IV. DATA & EXPERIMENT

The data for this project was retrieved from data.mendeley.com. The data consisted of 2298 records. These records had 26 descriptive features and an associated image with them. The features of the data were as follows patient's age, skin lesion location, Fitzpatrick skin type, and skin lesion diameter. The data had missing values and other issues which are discussed below.

A. Data Preprocessing

Data preprocessing is a necessary step for creating an accurate predictive model. It helps clean data and remove any unnecessary information from the data. The data after preprocessing becomes more easy to interpret and understand. There were several issues with the data that were meant to be addressed.

- Missing Values: There were 10484 missing values under the entire metadata data frame. The missing values in the dataset would contribute to errors in the predictions made by the machine learning model. For example, to demonstrate the missing values in the dataset, the smoke feature out of 24 features solely had 804 missing values. This shows us the importance to combat this issue hence we use the mean replacement method to fill the missing values. The mean of the corresponding columns was calculated and replaced with the missing records in the columns.

- Converting String to Integers: Another issue in the dataset was the problem of having string values of columns. For example, the drink feature had two values True or False, this would cause an issue since the machine learning algorithms would just see these values as a normal string. This would result in information loss if proceeded without addressing this issue. As a result, there were 12 other features that needed a type conversion from True to 1(int) and False to 0(int). This way the machine learning algorithms would make use of this information to improve the overall predicting accuracy.

- One-hot encoding for Multiple values: Another issue with the data was the presence of multiple values for various columns. For example, the descriptive feature region had different categorical values such as nose, scalp, thigh, forearm, hand, chest, and many more. The issue with this format was that it was too difficult for the machine learning algorithms to learn all these different categorical values for various columns. The solution for this was a one-hot encoding with pandas.get_dummies() for various categorical values, this way a new column is created with 1s and 0s indicating the presence of that particular value in the original column. Although it increased the number of descriptive features from 24 to 60 but this way the algorithm was able to learn from this information and make more accurate predictions.

- Target Features Label Encoder: There are six different target features values in the dataset as follows (a) BCC, (b) SCC, (c) MEL, (d) ACK, (e) NEV and (f) SEK. The issue with this is the difficulty for the machine learning algorithm to learn from the various string values. Hence the project uses the LabelEncoder method to encode different levels of categorical values of the target feature into numeric values, making it easier for the algorithm to learn from the numeric values.

V. Conclusion

In conclusion the new model proposed which uses the image data and text data, is definitely a good starting point for future enhancements in the accuracy of the existing model. This project implements a basic architecture to use image and textual data for skin cancer classification. The model makes a prediction with an accuracy of 68.7%. Some of the areas that were not touched upon was the image enhancement which includes removing noise from the images common techniques could be contrast enhancement, linear contrast adjustment etc. These next steps can help improve the overall accuracy of the model and it looks promising.

Acknowledgment

I would like to thank Dr. Jiashu Zhao for continued support and guidance throughout the process of development of this project.

References

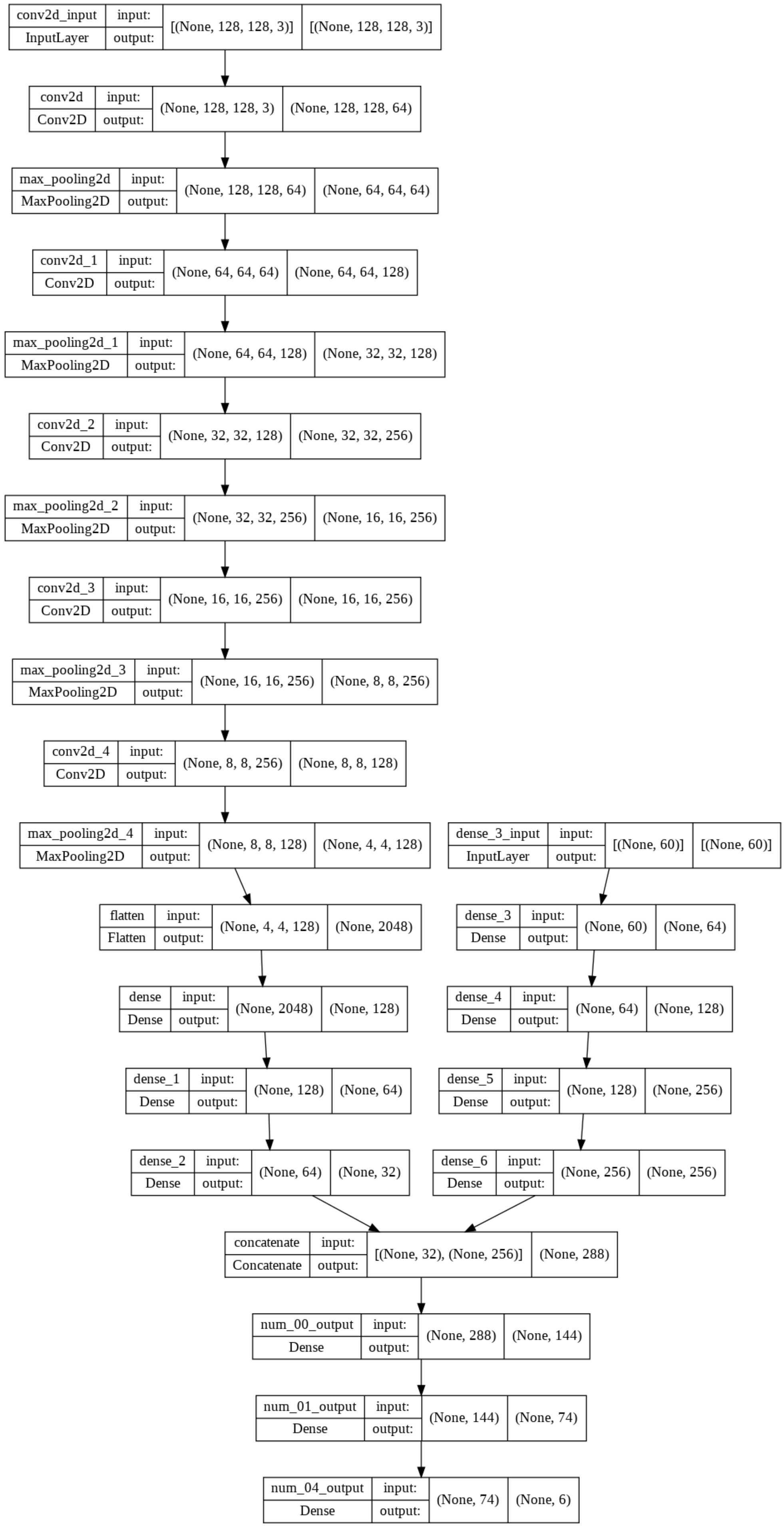
[1] "The Skin Cancer Foundation official website," The Skin Cancer Foundation, 04- Mar-2022. [Online]. Available: https://www.skincancer.org/. [Accessed: 17- Apr-2022].

[2] "The surgeon general's call to action to Prevent Skin Cancer- ncbi.nlm.nih.gov." [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK247 172/. [Accessed: 18- Apr-2022].

[3] M. K. Santos, J. R. Ferreira Júnior, D. T. Wada, A. P. M. Tenório, M. H. N. Barbosa, and P. M. de A. Marques, "Artificial Intelligence, Machine Learning, computer-aided diagnosis, and radiomics: Advances in imaging towards to Precision Medicine," Radiologia brasileira, 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7007049/. [Accessed: 17-Apr-2022].

[4] [4] Mahamudul Hasan Department of Computer Science and Engineering, M. Hasan, Surajit Das Barman Department of Computer Science and Engineering, S. D. Barman, S. Islam, and A. W. Reza, "Skin cancer detection using convolutional neural network: Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence," ACM Other conferences, 01- Apr-2019. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/3330482.33 30525. [Accessed: 17- Apr-2022]..

[5] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, "Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using MobileNet," SpringerLink, 01-Jan-1970. [Online].Available: https://link.springer.com/chapter/10.1007/978- 981-15-3383-9_15. [Accessed: 17-Apr-2022].


conv2d_input | InputLayer | input: [(None, 128, 128, 3)] | output: [(None, 128, 128, 3)]
conv2d | Conv2D | input: (None, 128, 128, 3) | output: (None, 128, 128, 64)
max_pooling2d | MaxPooling2D | input: (None, 128, 128, 64) | output: (None, 64, 64, 64)
conv2d_1 | Conv2D | input: (None, 64, 64, 64) | output: (None, 64, 64, 128)
max_pooling2d_1 | MaxPooling2D | input: (None, 64, 64, 128) | output: (None, 32, 32, 128)
conv2d_2 | Conv2D | input: (None, 32, 32, 128) | output: (None, 32, 32, 256)
max_pooling2d_2 | MaxPooling2D | input: (None, 32, 32, 256) | output: (None, 16, 16, 256)
conv2d_3 | Conv2D | input: (None, 16, 16, 256) | output: (None, 16, 16, 256)
max_pooling2d_3 | MaxPooling2D | input: (None, 16, 16, 256) | output: (None, 8, 8, 256)
conv2d_4 | Conv2D | input: (None, 8, 8, 256) | output: (None, 8, 8, 128)
max_pooling2d_4 | MaxPooling2D | input: (None, 8, 8, 128) | output: (None, 4, 4, 128)
flatten | Flatten | input: (None, 4, 4, 128) | output: (None, 2048)
dense | Dense | input: (None, 2048) | output: (None, 128)
dense_1 | Dense | input: (None, 128) | output: (None, 64)
dense_2 | Dense | input: (None, 64) | output: (None, 32)
dense_3_input | InputLayer | input: [(None, 60)] | output: [(None, 60)]
dense_3 | Dense | input: (None, 60) | output: (None, 64)
dense_4 | Dense | input: (None, 64) | output: (None, 128)
dense_5 | Dense | input: (None, 128) | output: (None, 256)
dense_6 | Dense | input: (None, 256) | output: (None, 256)
concatenate | Concatenate | input: [(None, 32), (None, 256)] | output: (None, 288)
num_00_output | Dense | input: (None, 288) | output: (None, 144)
num_01_output | Dense | input: (None, 144) | output: (None, 74)
num_04_output | Dense | input: (None, 74) | output: (None, 6)