Applied Machine Learning Coursework

Busra Yavuz
Applied Machine Learning
001190624

***Abstract*—Skin cancer is the most common type of cancer in the world today. For symptomatic patients, it is important to detect cancer as early as possible and use the right treatment. When cancer treatment is delayed or unavailable, there are greater problems associated with treatment, a lower chance of survival, and higher costs of care. Dermatoscopic images are the most important data for early detection of skin cancer. Computer-aided applications involving artificial intelligence models are a useful tool in assisting specialists in diagnosing skin cancer from dermatoscopic images. In this study, Convolutional Neural Networks (CNN) was used to detect skin cancer from dermatoscopic images. Image identification, object detection, and segmentation are among of the most common image analysis tasks that CNN is employed for. The data set exhibited a significant class imbalance, which was addressed using an equal train-validation-test split across all classes and direct class loss weighting during model training. Four distinct CNN models were used to assess and forecast the different types of skin lesions as well as whether they are benign or malignant. The models used are VGG16, RestNet152V2, and Basic CNN. ResNet152V2, the model used to detect whether skin lesions are benign or malignant. ResNet152V2, the model used to detect whether skin lesions are benign or malignant, gave a better accuracy than Basic CNN with 85.37% accuracy. The VGG16 model, which was used for the skin lesion classes created for fine-grained diagnosis for skin lesions, it was expected to give a noticeably better result than the Basic CNN model. However, a satisfactory result could not be obtained due to the insufficient performance of the computer used.**

I. Introduction and related work

Skin lesion analysis is important in medicine. It collects healthcare data from patients through hospital visits, medical research. The reason for collecting this data is that it allows to discover and analyze statistics about the disease in the diagnosis and treatment of diseases. Accurate analysis of the obtained data helps in the early diagnosis of diseases. Accurate analysis of data, such as malignant pigment lesions, is useful in the early diagnosis of skin cancer, so that some measures can be taken to destroy the cancer and the risk of death can be reduced. With semi- or fully-automatic computer-aided diagnostic systems, analysis can be done through machine learning, which will create a framework to help doctors communicate objectively, help reduce and decrease mortality rates, increase clinical reliability. In this way, diseases can be identified more easily and costs can be reduced further. A machine learning algorithm that can classify both malignant and benign pigmented skin lesions is a step in the right direction towards delivering these benefits.

In this study, Convolutional Neural Networks (CNN) were used to detect cancerous skin lesions as early as possible and to classify skin lesions in dermoscopic images in the most appropriate way. CNNs are fully connected feed forward neural networks. CNNs are very effective in reducing the number of parameters without losing on the quality of models. [2] Images have high dimensionality . In order to perform fine-granied analysis, 2 different CNN models were used for seven pigment lesion classes. The models used for this are Basic CNN architecture and ResNet152V2. Basic CNN model and VGG16 models were used to determine whether the skin lesion was malignant or benign.Information on these models can be found in IV. The models were compared after training the data using the metrics in section V. The data set used in this study is "The HAM10000 data set is a large collection of multi-source dermoscopic images of common pigmented skin lesions" and the information about the data set is discussed in more detail in section III. [5] The data set has large class imbalances as can be seen in section III.

A similar work was carried out by Rifat Edizkan in February 2021 in order to detect the skin cancer in a timely and accurate manner and to increase the probability of success of the treatment. Computer aided diagnostic applications were used for the classification of skin cancer. A two-class dataset consisting of dermoscopic images from the ISIC archive was used. Classification of benign and malignant skin cancer has been made and it is aimed to increase the success rate with the correct early diagnosis. For this, color sharpening, edge detection and noise reduction preprocesses were applied to dermoscopic images. Inception V2 model was used in the study. As a result of the study; With the preprocessing step, the success rate was increased by 3.33 points and the accuracy rate was obtained as 88.66%. [6]

II. Ethical discussion

It provides the ability to analyze comprehensive patient information in greater detail to monitor and distinguish between sick and relatively healthy people, which can lead to a better understanding of biological indicators that may indicate changes in health. Skin lesion detection can make it difficult to use healthcare information in clinical decision making, but existing difficulties can be minimized with technologies such as machine learning. With the study, the laboratory and the database and approaches discussed are important in order to increase the interoperability of public health systems and to address social and ethical issues with a balance of protecting

the privacy of health data. The use of machine learning for data analysis and management in this study, and the classification of data to optimize decision making, may provide convenience in the field of medicine. It can enable to provide real-time healthcare at lower costs.

III. Dataset Preparation

The data set used in this study is The HAM10000 data set, a large collection of multi-source dermoscopic images of common pigmented skin lesions. This includes 5269 dermoscopic skin pigmented lesion images. 1023 of these images were labeled as benign tumors and 4246 as malignant tumors. All images must be scaled to a constant size before being fed into the CNN since neural networks only accept inputs of the same size. The data set contains 4 attributes, each image:

- An Image ID [image id]
- The Localization of the Skin Lesion [localization]
- A Diagnostic Skin Lesion Category [cell_type]
- A binary data is_benign [is_benign]

The data set was first statistically analysed. Figure 1 shows the descriptive statistics of the data set contents.

	image_id	cell_type	localization
count	5369	5269	5357
unique	5369	7	15
top	ISIC_0029908	Melanocytic nevi	back
freq	1	3525	1193

is_benign	
count	5269.000000
mean	0.805846
std	0.395586
min	0.000000
25%	1.000000
50%	1.000000
75%	1.000000
max	1.000000

Figure 1 Descriptive statistics of the data set

As can be seen how there was a unique image id for each entry. This indicated that there were not duplicate images for the same image_id . Additionally, cell_type was counted 5269 but image_id 5369 contains image. This shows that cell_type contains null data. Furthermore, cell type dominated the skin lesion categories, with a frequency of 3525 pictures out of 5269, indicating a class imbalance issue in the data set. Figure 4 shows the distributions of 7 skin lesion categories. As seen in Figure 1, Melanocytic nevi represent the majority of the data set and dermatofibroma has the least representation. There is a class imbalance problem as can be clearly seen from looking at the graph.

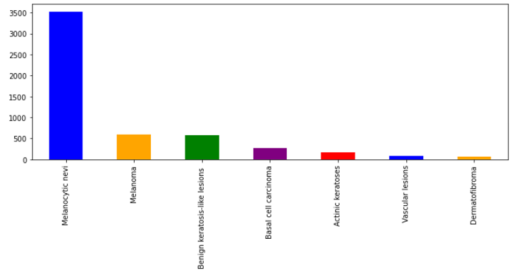

Figure 2 Skin Lesion Categories Distribution

The data set contains 7 different skin lesion classes. For this reason, there are 7 skin lesion categories. The categories are as follows:

- Actinic Keratoses:The Dataset contains 167 images of Actinic Keratoses.
- Benign Keratosis-like Lesions: The Dataset contains 275 images of Benign Keratosis-like Lesions.
- Melanocytic Nevi : The Dataset contains 3525 images of Melanocytic Nevi.
- Basal Cell Carcinoma : The Dataset contains 275 images of Basal Cell Carcinoma.
- Dermatofibroma: The Dataset contains 59 images of Dermatofibroma.
- Melanoma: The Dataset contains 587 images of Melanoma.
- Vascular Lesions:. The Dataset contains 77 images of Vascular Lesions.

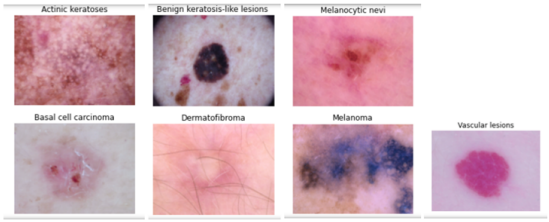

Figure 3 Image Samples for the 7 skin lesion categories

The data set's localization distribution is depicted in Figure 4. It's clear that the trunk, lower extremities, and back are all highly affected skin cancer locations.

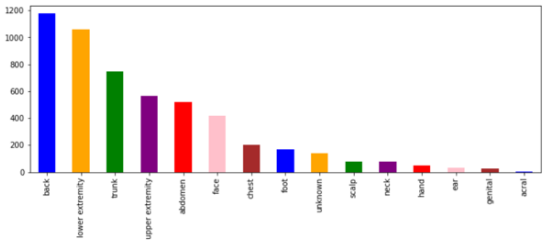

Figure 4 Skin Lesion Categories Distribution

Figure 5 shows the benign and malignant distribution of skin lesions in the data set. The malignant rate appears to be noticeably high.

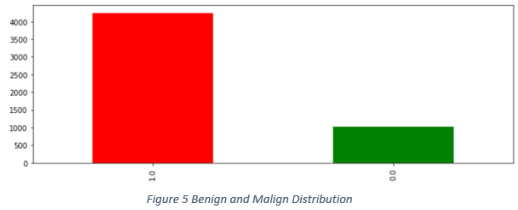

Figure 5 Benign and Malign Distribution

A. Data Editing and Cleansing

Two dictionaries were created separately for the data set containing the images and image names, and for the metadata containing the skin lesion categories code. The reason for this is to match the data set containing the images with the skin lesion codes in the metadata with the category full name. Only one image had to be kept for each image id and 4 duplicate values were detected. Duplicate image IDs have been removed. Also cell_type and is_benign contained null values. Null values are removed to provide a more reasonable estimation. For each skin lesion group, a unique number code was developed to help with future forecasts. Images have been

resized. The images have been resized so that the height and width of the images are not a problem when training the CNN model. Image size set to 150 by 112.5 pixels. One hot encoding generates new (binary) columns, one for each potential value from the original data. And this was provided seven unique label integer values with one hot encoding. It nullified any natural integer ordering assumptions obtained by the CNN algorithm.

B. Data Splitting

After the data cleaning, feature and target splitting was done. The feature is flattened image lists and is the only hot encoding created for target categories. Data was split into 70:10:20 for each class separately. The reason for choosing these values is to ensure that there are enough samples for each class and because of the imbalance in the data set. The training is divided into 70-Verification 10-Test 20.

C. Feature Normalization

Then the image feature was normalized. Normalization involves re-scaling the features such that they have the properties of a standard normal distribution with a mean of zero and a standard deviation of one. [3] Normalization for images was obtained by subtracting the values from the mean of the training and then dividing the result by the standard deviation of the training.

D. Data Augmentation

Image data augmentation was used to improve the performance and generalization ability of the model and to enlarge the training data set. Techniques such as panning, flipping, brightness adjustments and zooming were used. All of the original images were simply changed per epoch and then utilised for training. Because the model is trained on numerous versions of the same image, it is more resilient and accurate. The values are as follows: randomly zoomed by 10%, randomly channel shifted by 10%, randomly shifted vertically by 20%, randomly shifted horizontally by 20%, randomly rotated by 20%, Randomly sheared by 10%

E. Further Data Editing

To feed into the models, the flattened pictures were reformed as [width x height x depth]

IV. Methods

The models were implemented by using the Keras and Tensorflow libraries. Basic CNN and ResNet152V2 models were used in the study to determine whether the cancer is benign or malignant. In order to make a more detailed analysis, Basic CNN and VGG16 models were preferred for the analysis made over 7 categories.For categorical data, VGG16 is expected to outperform RestNet152V2.The structure of the models is as follows:

A. Basic CNN

This model includes 12 layers. Convolutional 2D layers are used because they make learning easier. Also instead of performing a convolution operation, the Maxpool layer selects the maximum values in the receiver fields of the input, saves the indices, and then a summarized output volume is produced. Relu was used because it does not activate all neurons at the same time. Neurons are disabled only if the output of the linear transform is less than 0. The padding type is called same because the output size is the same as the input size. With MaxPool 2D layer, the maximum value over an input window or each channel of the input is used to down sample the input along its spatial dimensions. The Dropout layer, which helps minimise overfitting, changes input units to 0 at random with a rate frequency at each step during training time. Flatten layer is used to flatten the input Layer 10 consisted of a FC input Dense layer with 128 units. In a multi-class problem, SoftMax assigns decimal probability to each class. The sum of their decimal probabilities must equal 1.0. This added restriction allows training to converge faster than it would without.

B. ResNet152V2

The ResNet152V2 neural network has 152 layers. ResNet152V2 is an artificial neural network that uses skip connections or shortcuts to leap over some layers to help develop a deeper neural network. It enables circumventing the problem of disappearing gradients while building deeper network layers. [1] These connections also aid the model by helping it to learn the identity functions, ensuring that the upper layer performs at least as well as the lower layer, if not better.

C. VGG16

The VGG16 model has 16 layers. It's a common picture classification technique that's simple to utilise with transfer learning. VGG16 refers to a set of 16 weighted layers. The most distinctive feature of VGG16 is that, rather than having a huge number of hyper-parameters, it is concentrated on having 3x3 filter convolution layers with stride 1 and always utilised the same padding and max pool layer of 2x2 filter stride 2. [4] VGG16 is a basic stack of convolutional and max-pooling layers followed by one another and eventually fully connected layers that would performed well for the purpose.

V. Experiments and evaluation

CNN models were used to get the best results for both steps. Different CNN models were used to compare the results for each step. Because CNN is fully connected feed-forward neural networks, it is very effective at giving the best results without compromising the quality of the models. The models used contained many parameters and due to the size of the data set, it took 30 hours to train the data with 25 epochs. Since the first step of the study was done on binary data, it took less time than the second step. It took a long time to get results with seven categories used for fine-grained inspection in the second step. It would be possible to get a better result by increasing

the number of epochs if the performance of the computer used allowed more processing. However, in order to obtain a better result from the data models, a more detailed and technical analysis was made. The following hyperparameters were used to examine the data in more detail: Optimizer (Adam)- Adam was simple to set up, and the default setup options worked well for the majority of situations, Loss Function- Categorical cross-entropy is a single-label categorization loss function. Only one category is acceptable for each data point in this case. Because a sample could only belong to one of the seven types of skin lesions for the second step, or it may be benign or malignant for the first step, this worked well, Epochs, Batch Size-The best results were obtained with a batch size of ten., Learning Rate. In addition, Class weight was used for the second step of the study. Since the melanocytic nevus category caused imbalance in the data, the data was adjusted to make the data more sensitive.

			Basic CNN	ResNet152V2	VGG16
Is_Benign					
	Accuracy	Train	80.72	85.379	
		Validation	80.42	84.460	
	Loss	Train	0.347	0.288	
		Validation	0.367	0.355	
	RMSE	Train	0.208	0.157	
		Validation	0.219	0.172	
	MSE	Train	0.118	0.105	
		Validation	0.129	0.118	
Categorical					
	Accuracy	Train	66.90		66.90
		Validation	66.90		66.90
	Loss	Train	1.25		1.20
		Validation	1.27		1.21
	RMSE	Train	0.20		0.2057
		Validation	0.20		0.207
	MSE	Train	0.0815		0.0792
		Validation	0.081		0.079

Figure 6 Models Training and Validation

Figure 6 shows the accuracy, loss, and errors of the validation and training data sets. The Loss function, which shows how far the estimated value is from the true value, gave a value close to 0 for the first step, as it should be. RMSE train and validation values had a good result with a value close to 0, which shows that the error value between the estimated values and the actual values is very low. MSE shows that with a value close to zero, the estimator performs well. For the categorical part, the high loss value shows that the estimated value is far from the real value. The reason why both models have the same accuracy for this step is that the model cannot bend well with 25 epochs due to insufficient computing resources. 50 epochs gave a better result for VGG16.

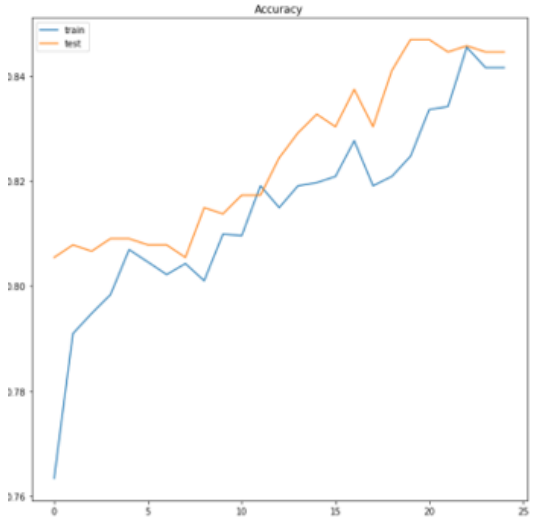

Figure 7 Training against epoch

Figure 7 shows the training of the RESTNET152V2 model against the epoch number. For the first step of this work, it shows the improving the accuracy of the model with each epoch while optimizing the function.

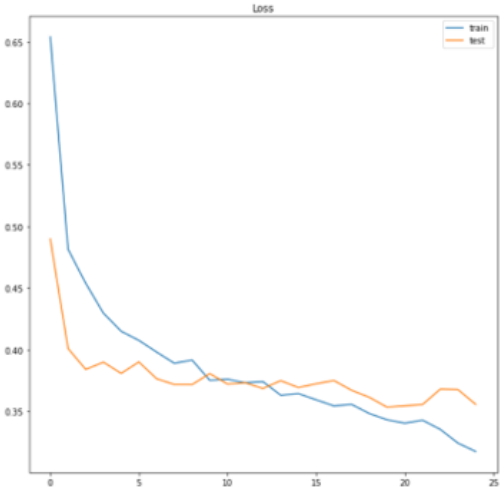

Figure 8 Training against epoch

Figure 8 shows the training of the RESTNET156V2 model against the epoch number. For the first step of this work, it shows the improving the lose of the model with each epoch while optimizing the function.

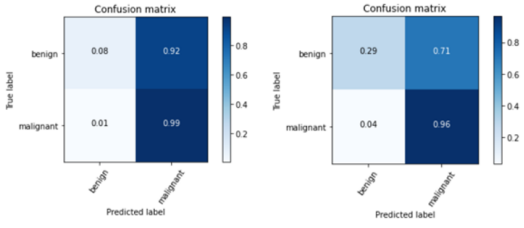

Figure 9 Basic CNN and RESTNET156V2 Confusion Matrices

Figure 9 shows the confusion matrices for the correct and predicted labels for each model when run on the test set.

The previously mentioned discussed metrics were used to accurately capture the values of true and false positives and negatives for each model.

	precision	recall	f1-score	support
benign	0.00	0.00	0.00	205
malignant	0.81	1.00	0.89	849
accuracy			0.80	1054
macro avg	0.40	0.50	0.45	1054
weighted avg	0.65	0.80	0.72	1054

	precision	recall	f1-score	support
benign	0.66	0.29	0.40	205
malignant	0.85	0.96	0.90	849
accuracy			0.83	1054
macro avg	0.75	0.63	0.65	1054
weighted avg	0.81	0.83	0.80	1054

Figure 10 Results of other performance metrics for Basic CNN and RESTNET152V2, respectively

Figure 10 shows the weighted testing sets recall, accuracy, precision, and f1-score linked with each model are listed. To begin with, all of the testing accuracies were equivalent to the validation accuracies, confirming that no model was overfitting or underfitting.

VI. Discussion and future work

The challenging part for this study was the failure to provide sufficient technical requirements to train the data. Working on binary data was fairly easy, but with better performing computing systems, the epoch number can be selected as 50 or higher to train the data better for skin lesion categories and achieve noticeable accuracy .However, using the Basic CNN model twice for binary data and categories, and training the data for performance comparison plus RESTNET152V2 for binary data and VGG16 for categorical data took more time than expected. In addition, it also reduced the performance of these models. Furthermore, the work may be enhanced by modifying the kind and number of layers, as well as experimenting with various class imbalance strategies to improve skin lesion category classifications for underrepresented classes.

VII. Conclusions

CNN was considered to be a more suitable model for this study, as it was thought that the data set containing the image and the data set containing the ids of images and skin lesion attributes of these images could be analysed in detail. Especially in the second step of the study, this feed-forward neural network could well reflect the performance difference between the models in order to make an appropriate analysis for the seven categories. Nevertheless, the computer resources used for the study were technically hindered. Since the analysis for binary data consisted of only two categories, an accuracy of 80.72% with the Basic CNN model and 85.37% for RestNet152V2 was obtained. However, in the study with seven categories, 25 epochs had to be selected due to computer performance, so enough epochs could not be used to train the data. The result was 66.90% accuracy for Basic CNN and VGG16. The codes took a total of 30 hours to run. In the study performed separately on another notebook for seven categories, 74.37% accuracy results of the epoch 50 and VGG16 model were obtained. However, since all the codes are required to be found and run on the same notebook, running both steps on the same notebook causes performance problems, so a satisfactory result could not be achieved with the VGG16 model, which normally performs well.

References

[1] Dina M Ibrahim and Nada M Elshennawy. Improving date fruit classification using cyclegan-generated dataset.

[2] Yanan Sun, Bing Xue, Mengjie Zhang, Gary G Yen, and Jiancheng Lv. Automatically designing cnn architectures using the genetic algorithm for image classification. *IEEE transactions on cybernetics*, 50(9):3840–3854, 2020.

[3] Dang NH Thanh, Uğur Erkan, VB Surya Prasath, Vivek Kumar, and Nguyen Ngoc Hien. A skin lesion segmentation method for dermoscopic images based on adaptive thresholding with normalization of color models. In *2019 6th International Conference on Electrical and Electronics Engineering (ICEEE)*, pages 116–120. IEEE, 2019.

[4] Dhananjay Theckedath and RR Sedamkar. Detecting affect states using vgg16, resnet50 and se-resnet50 networks. *SN Computer Science*, 1(2):1–7, 2020.

[5] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018.

[6] Feyza Yılmaz and Rifat Edizkan. Improvement of skin cancer detection performance using deep learning technique. In *2020 28th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE, 2020.