|  |  |  |
| --- | --- | --- |
| Date | : |  |

|  |  |  |
| --- | --- | --- |
| **TITLE** | **:** | **An Ensemble of Convolutional Neural Networks for Dermoscopic Image Classification** |

|  |  |  |
| --- | --- | --- |
| **Name of Student** | **Student No.** | **Program of Study** |
| **Camila, Dexter B.** | **2015112278** | **BSCS** |
| **Reyes, Alexander** | **2010108525** | **BSCS** |
| **Vega, Asaph F.** | **2015111865** | **BSCS** |

**IMPLEMENTATION OF THE RECOMMENDED REVISIONS**

| **Nature of Revision** | **Actions Made** | **Page Reflected** Manuscript |
| --- | --- | --- |
| 1. What is the result of the validation by the dermatologist as discussed in the methodology? | * 1. Validation:   A final model evaluation was performed by comparing the predictions with the ground truth data and results reviewed by a licensed dermatologist, Dra. Patricia Tinio of Makati Medical Center. | pp. 38 |
| 1. Please illustrate the conceptual framework you adapted | 2.1 Conceptual framework - inserted and specified as follows:  The conceptual framework of this study was derived from a few prominent, yet relative studies of Ronneberger [7] and Li et al. [26]. In the former study, the U-Net architecture was used to segment biomedical images for cell tracking and yielded outstanding results. This study utilized the same techniques in segmenting skin lesions from dermoscopic images. The latter study encompassed the process of classifying skin lesions efficiently using the VGG16 + ResNet50 architecture.  The training methodology depicted above was inspired by the studies of the two aforementioned researchers. Using the popular HAM10000 7-class skin lesion dataset, a map was created to easily identify each image skin type during the development process and to simplify the preprocessing methods. This was followed by a resampling technique called downsampling which removed some of the images belonging majority class to minimize the disparity amongst the classes and therefore reducing the possibility of overfitting.  2.2 Illustration – Graph of Conceptual Framework inserted | pp. 27 & 28 |
| 1. Provide detailed discussion of how you have done your experiment, from data gathering to validation (with illustration – if applicable) | * 1. Illustration of training process inserted   3.2 Description of training methodology elaborated here:  … The training methodology depicted above was inspired by the studies of the two aforementioned researchers. Using the popular HAM10000 7-class skin lesion dataset, a map was created to easily identify each image skin type during the development process and to simplify the preprocessing methods. This was followed by a resampling technique called downsampling which removed some of the images belonging majority class to minimize the disparity amongst the classes and therefore reducing the possibility of overfitting. After downsampling, the images were then uniformly resized (**Figure 4.2**)   * 1. VGG16 and ResNet50 results of training and address overfitting issues:   The results after training the classification models still showed signs of overfitting as the validation accuracy rates digressed.  … In order to combat this, class weights were computed and passed into the bias parameter of each classification model. This drastically improved the validation accuracy rates as well as the loss rates.   * 1. U-Net model adjustments and results:   The network used followed the same network structure as described in the study of Ronneberger [7]. One significant modification was made, however, as the dimensions of the input layer was not compatible with the HAM10000 dataset and the output of the classifier models. The original study had also set the epochs to 500 which, as with the classifier models, was not feasible considering the hardware limitations of this study. Epochs again were set to 100 for the same reasons. Accuracy and loss rate results were more than acceptable and only a few points less than what Ronneberger [7] had achieved.   * 1. Data gathering:   Datasets from preceding research [23] were gathered in order to maintain consistency amongst the results across the studies mentioned in this paper (ISIC 2018, HAM10000). Images in the “Human Against Machine with 10000 training images” dataset (HAM10000) include a representative collection of all-important diagnostic categories in the realm of pigmented lesions   * 1. Validation:   In order to validate the segmentation model, the ISIC 2018 dataset was split between training data and validation data (70% and 30% respectively). The overview of the complete training process was illustrated in Figure 4.1. The K-fold cross-validation resampling procedure was implemented in the training process of the classification models (VGG-16 + ResNet50) using the HAM10000 dataset.  … A final model evaluation was performed by comparing the predictions with the ground truth data and results reviewed by a licensed dermatologist, Dra. Patricia Tinio of Makati Medical Center. | pp. 28, 30, 31, 33, 38 |
| 1. Clearly discuss the results of your testing based from the method you applied | 4.1 Add discussion of test results:  The table above represents the classification results of both unaltered and segmented images. The VGG16 and ResNet50 ensemble did not produce desirable results. There are inconsistencies amongst the predictions of all classes in both classification methods and the confidence of the predictions are quite low. In only one instance did the segmented image produced a higher percentage rate; not to mention predicting the correct class.  … According to Ronneburger [7], specificity is often considered the most important metric in a medial setting for early diagnosis. By missing a false negative (a true malignant case), the model would fail in the early diagnosis. One could surmise that it would be better to raise a false positive than create a false negative. | pp. 43 - 44 |
| 1. Please discuss the result of your validation scientifically. Did you make a program (a proof of concept) to validate the result of your experiment? | 5.1 Insert validation result tables  5.2 Add discussion of validation:  The VGG16 and ResNet50 ensemble did not produce desirable results. There are inconsistencies amongst the predictions of all classes in both classification methods and the confidence of the predictions are quite low. In only one instance did the segmented image produced a higher percentage rate; not to mention predicting the correct class.  … Most of the results show a decrease in accuracy when comparing unaltered with segmented images. Referring to the Melanocytic Nevi image, accuracy decreases while loss increases (**Figure 5.5**). | pp. 44 |
| 1. Place your detailed screenshot/results in appendix and leave the most significant only in your main documentation, with discussion | 6.1 Moved screenshots to appendix | pp. 46 |
| 1. Present a table to summarize your obtained values.  Interpret your results afterwards | 7.1 Add table illustrating results and discussion:  The table above represents the classification results of both unaltered and segmented images. The VGG16 and ResNet50 ensemble did not produce desirable results. There are inconsistencies amongst the predictions of all classes in both classification methods and the confidence of the predictions are quite low. In only one instance did the segmented image produced a higher percentage rate; not to mention predicting the correct class. | pp. 43 |
| 1. Present confusion matrix or any applicable way to illustrate your findings | 8.1 Add Confusion matrix and discussion:  The confusion matrix above reflects the class imbalance in the dataset. Unsurprisingly, the Melanocytic Nevi majority class is the most accurately predicted class due to its overwhelming presence in the dataset even after apply downsampling. The bar graph below represents the percentages of incorrect predictions derived from the computations of the confusion matrix. It reaffirms the Melanocytic Nevi class is the most accurately predicted class. Unfortunately, it also indicates that the remaining classes are only predicted correctly between 12% and 57% of the time. | pp. 42 - 43 |
| 1. These should have mathematical application...how as these drived | 9.1 Show mathematical formulas applied when deriving results:  The sensitivity and precision metrics were calculated using the following formulas:  **Specificity = TN/FP+TN**  **Precision = TP/TP+FP** | pp. 44 |