IRDS MiniProject: Identifying Malaria Parasites from Images

Mica Teo s1459898

Abstract

Microscopy examination has been the pillar of current malaria diagnosis, being the recommended procedure when its quality can be maintained. However, rapid, accurate, accessible diagnostic tools are increasingly required. In this paper, the performance of three methods of classification are evaluated: K-Nearest Neighbour (KNN), Extremely Randomized Tree Classifier (ERTC), and Convolutional Neural Network (CNN) on diagnosis of malaria in thick blood smears. The CNN provides a good performance in malaria detection task.

1. Introduction

The gold standard test for malaria is the hundred-year-old method of preparing a blood smear on a glass slide, followed by staining it, and the examining the sample under a microscope for the parasite genus plasmodium. Although malaria is most entrenched in poor, tropical and subtropical areas of the world such as Uganda, however, the access to this standard test is often limited and resulting in misdiagnosis. Accurate diagnosis is important since false negatives could result in deaths, whereas false positive could cause drug resistance, and country economic burden by buying unnecessary drug. In addition to the limited access of gold standard test, lack of funding to hire an experienced lab technician to conduct the diagnosis test also result in misdiagnosis. For instance, recent survey carried out in Uganda found 50% of health centres have microscopes, but only 17% have trained laboratory technicians for malaria diagnosis. Moreover, a nationawide study in Ghana, (Petti et al., 2006), found only 1.72 microscopes available per 100,000 population. Lab technician spent excessive time to give a confident diagnosis. The proposed solution is to capture and process microscopy images by using phone. The aim is to process the images from the microscope in order to focus the technician's attention on only the objects within those images that most resemble plasmodium. For this, a different threshold with greater sensitivity would be appropriate.

2. Data Preparation

2703 blood smear images downloaded from (A. Quinn et al., 2014). Each of the downloaded blood smear images was annotated with bounding boxes around each malaria parasite using labelling software developed for the PASCAL Visual Object Classes challenge (Everingham et al., 2015). In this way the coordinates of 50 255 parasites were recorded within the set of captured images. Each images was downsampled by factor of 2 and split up into overlapping patches. Each patch was assigned with a label of 0 or 1. Positive patches are those containing plasmodium whereas negative patches are those absence of plasmodium as shown in Figure 1. The plasmodium detection task is a classification problem.

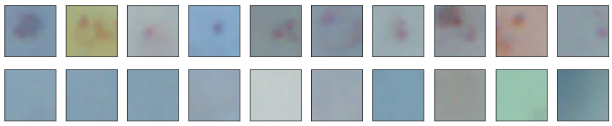

Figure 1. **Top**: A random selection of positive patches. **Bottom**: A random selection of negative patches.

The data is imbalanced. There are typically many potential negative patches in each image since most of each image does not contain malaria parasites. Thus, disproportionately large number of negative patches are cut down, and only the proportion 0.1 of negative patches are stored. Negative patches were randomly discarded, and the upper limit of the number of negative patches was at most 100 times the number of positive patches. Besides, seven extra positive examples were created from each positive-labelled patches by applying different combinations of rotation and flipping.

3. Feature Engineering

For content-based image, the representation based on an array of pixels is not appropriate as we want to be able to act on objects, to selectively encode areas of interest, that is to detect malaria parasite. Therefore, the notion of object is essential. The raw form of the pixel data in these image patches is not directly very useful for classification image patches. A representation which is invariant with respect to rotation, translation and constant offsets in intensity is required. Feature engineering is therefore a crucial step in developing
tures are used: connected component features and moment features. Before feature engineering, each colour patch was converted to grayscale.

3.1. Connected component features:

Connected operators are filtering tools that act by merging elementary regions called flat zones (Salembier & Wilkinson, 2009). Connecting operators cannot create new con-

Type of Data	**Training Data**		**Testing Data**		**% Positive**
	Positive	Negative	Positive	Negative	
Non-augmented	36121	1168201	12134	388773	3.0 %
Augmented	288968	1169434	97072	388571	19.9 %

Table 1. The proportion of positive and negative patches in training and testing data before and after augmenting.

tours nor modify their position. Therefore, they have very good contour-preservation properties and are capable of both low-level filtering and higher level object recognition (Salembier & Serra, 1995).

Morphological features :

- Cityblock perimeter
- Large perimeter
- Moment of Inertia
- Elongation: Inertia/Area^2
- Jaggedness: (Area $\times$ Perimeter^2) / ($8\pi^2\times$ Inertia)
- Maximum λ: Maximum child gray level - current gray level

are computed for each grayscale image patch at the 25th, 50th, 75th percentiles, minimum and maximum for all the shapes in a max-tree. Thus, a 30-dimensional feature vector for each image patch are obtained.

3.2. Moment features:

Moment statistics (Teh & Chin, 1988) :

- Moment m_{00}
- Central moments $\mu_{11}, \mu_{20}, \mu_{02}$
- Hu moments h_0, h_1, h_2

are computed for each grayscale image patch by thresholding the grayscale image patch at five different levels between the minimum and maximum pixel value. Thus, an additional of 35-dimensional feature vector for each image patch are obtained.

4. Explaratory Data Analysis

By adjusting the classification threshold, different threshold, different trade-offs are found between false positives and false negatives. Figure 2 shows the sample detection output on part of a test image by setting different threshold. The threshold is set as 0.25, 0.5 and 1.0, increased from left to right. If the threshold is set too high, the false positive fraction will decrease with increased specificity but on the other hand the true positive fraction and sensitivity will decrease. On the other hand, when a lower threshold value is selected, the true positive fraction and sensitivity will increase. The false positive fraction will also increase, and thus the true negative fraction and specificity decreased.

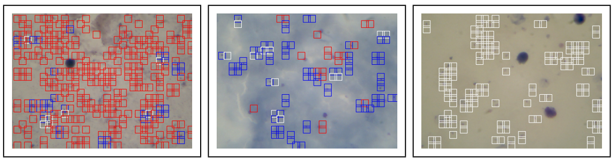

Figure 2. Sample detection output on the test image. **Blue** squares indicate patches which were correctly classified as containing parasites. **Red** squares indicate false positive and **White** squares indicates false negatives.

5. Classification Methods

The labeled data was split into training and testing subsets. Each classifier listed below was trained on 75% of the labeled data (2027 images, containing 36 121 patched annotated as containing parasites), and tested on remaining 25% of the labeled data (676 images, containing 12 134 patched annotated as containing parasites). The number of positive and negative patches in training data and testing data before and after augmenting are tabulated in Table 1. After augmenting, the percentage of patches containing malaria parasites increase from 3% to 19.9%.

Cross validation with 5-fold split is used on the training data set, to perform a grid search for the best value of hyperparameter for each chosen model. Then, the optimal values, the hyperparameters that yield the highest accuracy, is used to train a new classifier model on the entire training set.

5.1. K-nearest neighbour (KNN)

The K-nearest neighbour (KNN) categorizes objects based on the classes of their nearest neighbors in the dataset. KNN predictions assume that objects near each other are similar. It was used as a simple algorithm to establish benchmark learning rule. It is used to learn a mapping between features and the patch labels. However,the memory usage of the trained model is big, and the prediction speed of the trained model is pretty slow.

8-KNN is used to train the data, where the optimal value of parameter, number of neighbor, $k = 8$ was found by using 5-fold cross-validation. It selects the closest 8 training points and predicts the most frequent class in that subset, and it is less sensitive than 1-NN to outliers.

5.2. Extremely Randomized Tree Classifier (ERTC)

The Extremely Randomized Trees classifier (ERTC), (Geurts et al., 2006) was used to learn a mapping between features and the patch labels. The random decision forest is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees. An ensemble of 250 trees, with a maximum tree depth of 5 was used. One of the great advantages of decision trees is their interpretability. The rules learnt for classification are easy for a person to follow, unlike the opaque "black box" of many other methods, such as neural networks.

5.3. Convolutional Neuron Network (CNNs)

Convolutional Neural Networks (CNNs) became more and more important in the last few years, especially in terms of image and computer vision. It gradually dominates various computer vision tasks. A powerful aspect of using convolutional networks for image classification is their ability to incorporate information at different scales.

CNNs use three basic ideas: local receptive fields, shared weights, and pooling layer. In addition to the convolutional layers, CNNs contain pooling layers which is an important concept of CNNs purposed for non-linear down-sampling, comes after convolutional layers. This layer serves the purpose to condense information from the previous layer. For instance, max-pooling extracts the most important features like edges.

The Convolutional Neuron Network(CNN) was used here to learn a mapping between features and the patch labels. The CNN model architecture of contains two standard convolutional layers, two max-pooling layers which are placed after each convolutional layers, and one fully connected layers with 500 hidden neurons. The first convolutional layer filters input patches with seven 3×3 kernels with a stride of 1 pixel, which means the distance between the receptive field centers of neighbouring neurons in a kernel map is one. The first max-pooling layers performs a striding over 2×2 neigbourhoods. The second convolutional layer takes the output of the first max-pooling layer as input and filters it with 12 kernels of same size. Same principles apply to second max-pooling layers. Lastly, the outputs are fed to the fully-connected layer with 2 output units, where each of the 2 numbers correspond to positive or negative patch.

The activation function is the Rectified Linear Unit (RELU), except for the last layer where softmax is used. The cross-entropy is used as the cost function in the model. The model is trained with ADAM Weight decay, an initial learning rate of $1.0e-05$.

6. Results and Discussions

This paper addresses three questions: (1) Does the performance of classifier improve when the augmented data set is used? (2) Does the performance of classifier improve when instead of pixels, a set of morphological and moment statistics are used as shape features and applied to classifier? (3) In paper (Quinn et al., 2016), it was mentioned that high accuracy of malaria detection task under CNN likely due to having the largest training set. Thus, I would like to investigate whether the accuracy will stay high if the training set is reduced by 25% or 50%.

The receiver operating characteristic curve (ROC) and the precision-recall curve (PR) are used to evaluate the performance of each classifier. The areas under receiver operating characteristic curve shows how well a classifier can distinguish between positive and negative patches, while the precision-recall curve shows different trade-offs possible between increasing sensitivity and decreasing the false alarm rate.

6.1. K-nearest neighbour and Extremely Randomized Tree Classifier (ERTC)

From Figure 3, it looks like the performance of classifier would keep going up if more training data is used to train the model. However, the percentage of increased Area Under Curve (AUC) gradually diminish. Thus, we might consider to use smaller set of training data set if we take into consideration of computation time.

The model generalisation appears to improve as parameter k under k-nearest neighbour classifier increases but only up to a point as shown in Figure 4. If we select k to be too high, then we will find that all the instances will be classified in the most probable class, and if we set it too low we will find that the decision boundary is unstable. So one way to pick k is try different k, and pick the k that performs best on a validation set.

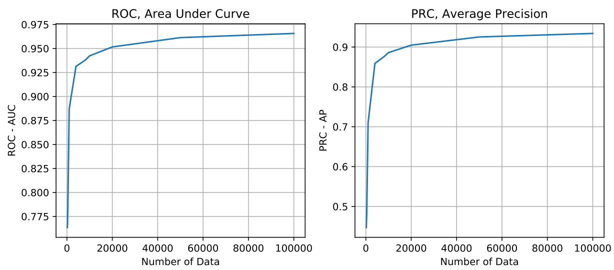

Figure 3. Plot of Area Under Curve (AUC) against number of data.

Receiver Operating Characteristics and Precision-Recall curves are shown for each case in Figure 5 and Figure 6. In overall, the performance of both classifiers improve when using the augmented data as tabulate in Table 2. PR curves are better than ROC curves in evaluating the performance due to the high class imbalance in the data set. In the case of using pixel as features, for KNN, the average precision increases significantly from 0.47 to 0.93 whereas for ERTC, the average precision greatly increases from 0.65 to 0.97. In the case of using shape features which are extracted from each patch, for KNN, the average precision substantially increases from 0.33 to 0.77 whereas for ERTC, the average

Methods	Data	Feature	ROC	AP
KNN	Non-augmented	Pixel	0.76	0.47
	Augmented	Pixel	0.97	0.93
	Non-augmented	Connected component features + Moment features	0.64	0.33
	Augmented	Connected component features + Moment features	0.95	0.77
ERTC	Non-augmented	Pixel	0.97	0.65
	Augmented	Pixel	0.99	0.97
	Non-augmented	Connected component features + Moment features	0.99	0.58
	Augmented	Connected component features + Moment features	0.99	0.97

Table 2. Area Under Curve (AUC) and Average Precision (AP) for classification methods: k-nearest neighbour (KNN) and Extremely Randomized Tree Classifier (ERTC) for using different cases of data and feature.

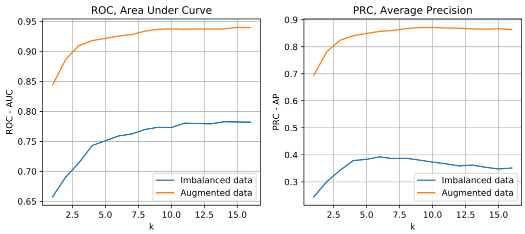

Figure 4. Plot of Area Under Curve (AUC) against parameter k for non-augmented and augmented data under k-nearest neighbour classifier.

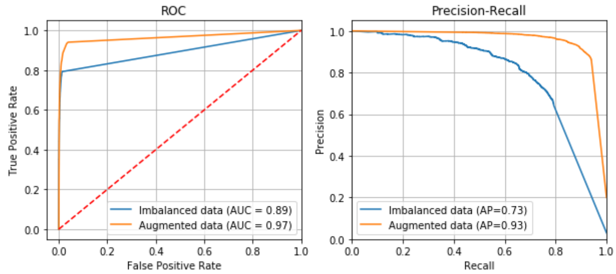

Figure 5. Receiver Operating Characteristics and Precision-Recall curves for test data, showing Area Under Curve(AUC) and Average Precision(AP) under k-nearest neighbour classifier.

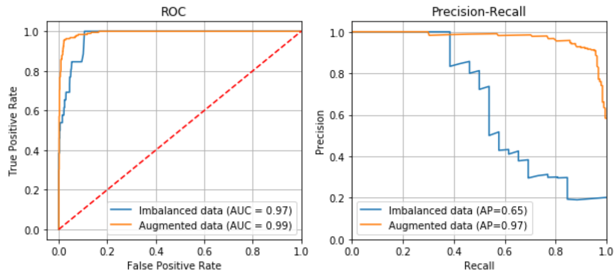

Figure 6. Receiver Operating Characteristics and Precision-Recall curves for test data, showing Area Under Curve(AUC) and Average Precision(AP) under Extremely Randomized Tree Classifier (ERTC).

precision rises significantly from 0.58 to 0.97. The high value of AUC indicates that the classifiers are effective at distinguishing positive and negative patches. The average precision for the model using either pixels or shape features is almost similar.

To have greater sensitivity, we might require recall 90 %, giving corresponding precision. For instance, under KNN with non-augmented data, we would expect the system to detect nine out of every ten parasites appearing in the images, but giving corresponding 30 % precision, that is 70 % of detections would be false alarms. On the other hand, under KNN with augmented data, we would expect the system to detect nine out of every ten parasites appearing in the images, but giving corresponding 90 % precision, that is 10 % of detections would be false alarms.

6.2. Convolutional Neuron Network

Table 3 shows the change of Area Under Curve (AUC) and Average Precision (AP) by using different percentage of train data to train CNNs. There is no much significant different if we decrease train data to around 25 %. Thus, the high accuracy of the performance of CNNs might not due to the large data set.

7. Conclusions

Fron th experiments results, nearest neighbour is not the most effective classification model for this problem as it is computationally expensive at testing time. For classification of images, the best types of models are those which take into account that the inputs are pixels which can be viewed on a grid, and do not just assume that every input dimension is completely independent of the others. This is one reason that convolutional networks are often effective for this sort of problem. When there is imbalances data, it would be better to augment the data by rotating or flipping the original images to increase the proportion of positive patches.

For future works, I would to investigate different types of filter on improving the performance of classification, for instance dilation or segmentation.

% of Train Data Used	ROC	AP
100 %	0.98	0.93
75 %	0.98	0.93
50 %	0.98	0.92
25 %	0.98	0.92
10 %	0.97	0.92
1 %	0.93	0.78
0.1 %	0.50	0.18

Table 3. Area Under Curve (AUC) and Average Precision (AP) under CNNs at different proportion of used train data.

References

A. Quinn, John, Andama, Alfred, Munabi, Ian, and Kiwanuka, Fred. Automated blood smear analysis for mobile malaria diagnosis, 09 2014.

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, January 2015.

Geurts, Pierre, Ernst, Damien, and Wehenkel, Louis. Extremely randomized trees. *Machine Learning*, 63 (1):3–42, Apr 2006. ISSN 1573-0565. doi: 10.1007/s10994-006-6226-1. URL https://doi.org/10.1007/s10994-006-6226-1.

Petti, Cathy A., Polage, Christopher R., Quinn, Thomas C., Ronald, Allan R., and Sande, Merle A. Laboratory medicine in africa: A barrier to effective health care. *Clinical Infectious Diseases*, 42(3):377–382, 2006. doi: 10.1086/499363. URL http://dx.doi.org/10.1086/499363.

Quinn, John A., Nakasi, Rose, Mugagga, Pius K. B., Byanyima, Patrick, Lubega, William, and Andama, Alfred. Deep convolutional neural networks for microscopy-based point of care diagnostics. *CoRR*, abs/1608.02989, 2016. URL http://arxiv.org/abs/1608.02989.

Salembier, P. and Serra, J. Flat zones filtering, connected operators, and filters by reconstruction. *IEEE Transactions on Image Processing*, 4(8):1153–1160, Aug 1995. ISSN 1057-7149. doi: 10.1109/83.403422.

Salembier, P. and Wilkinson, M.H.F. Connected operators: A review of region-based morphological image processing techniques. *IEEE Signal Processing Magazine*, 6:136–157, 2009. ISSN 1053-5888. doi: 10.1109/MSP.2009.934154.

Teh, C. and Chin, R. On image analysis by the methods of moments. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 10:496–513, 07 1988. ISSN 0162-8828. doi: 10.1109/34.3913. URL doi.ieeecomputersociety.org/10.1109/34.3913.