Internship report

Topic:

Learn Parasitology and Diagnose with Deep Learning

Mohsin Kaleem

IIT Kharagpur

1 Introduction

Machine learning methods have become quite popular in prediction tasks, especially because of the success of deep neural networks (DNN). From language and speech processing tasks to image recognition, DNN's have proved their worth. This has enabled us to automate many processes, be it in day-to-day life, in industries or in health sectors. However, these techniques are often treated as a black box and it becomes important to make use of additional attributes in the data to help avoid incorrect decisions and keep more informed decisions. One of the ways to account additional attributes in classification task is by incorporating a multi-label DNN. A multi-label DNN is a generalization of a multi-class DNN in which each observation sample belongs to multiple classes. Thus, besides the main prediction, if the data has labels for additional attributes, this can be utilized to gather additional information along with the main prediction. Multi-label DNN's have been applied for intelligent risk prediction (1), chronic disease prediction (2), audio tagging (3), facial attribute classification (4) and clinical document classification (5). In this study, a two-stage classification system is adopted. The first stage is the prediction of the species of the parasite from their images. For inspecting the areas in the images that contributed to the prediction of the respective species, Gradient-weighted Class Activation Mapping (Grad-CAM) was used. The second stage of the prediction is the prediction of attributes of the image. The attributes, contributing to the images were the Phylum, Class, Form and Sample of the parasite. In this stage, classification was performed

by a by a multi-label DNN. For using relations between the labels, a multi-head self-attention block was used and was compared with another model with same architecture but without the attention block (baseline model). The DNN incorporating multi-head self-attention block performed better than the baseline. Keeping in view the small sample size and a large number of labels, results show a promising direction in this area of study.

2 Method

The overall workflow is divided into the following sections: (i) Gathering data from the CDC website (6), (ii) Annotating the images, (iii) Pre-processing the data, (iv) Prediction of species, (v) Prediction of attributes.

2.1 Gathering data from CDC website

From the CDC website (6), the parasitology data was downloaded using web scraper. The data comprised images of the parasite as gathered from microscopic as well as macroscopic examinations. The organisms for which the images were downloaded are mentioned in table 1:

Ancylostoma sp.	Entamoeba histolytica	Plasmodium vivax
Ascaris lumbricoides	Enterobius vermicularis	Sarcocystis sp.
B. coli	Fasciola hepatica	Strongyloides stercoralis
Clonorchis sinensis	Giardia duodenalis	Strongyloides fuelleborni
Cryptosporidium sp.	Hymenolepis diminuta	Taenia solium
Cyclospora cayetanensis	Hymenolepis nana	Toxocara canis
Dibothriocephalus latus	Plasmodium falciparum	Toxoplasma gondii
Dicrocoelium dendriticum	Plasmodium knowlesi	Trichinella sp.
Echinococcus granulosus	Plasmodium malariae	Trichuris Trichuria
Echinococcus multilocularis	Plasmodium ovale	

Table 1: List of organisms for classification.

2.2 Annotating the image

The images were annotated with respect to their Phylum, Class, Species, Form and Sample. While the Phylum, Class and Species represented their respective taxonomic category, form represented the morphological state of the organism (for example, whether the image represented the parasite in egg or the adult form) and sample represented the part of the body or the environment from which the parasite was drawn (for example blood sample, tissue sample, etc.). Images were annotated manually. While downloading the images from the web scraper, other metadata such as phylum, class, species and names of the images were also downloaded simultaneously and saved in a csv file. Remaining metadata such as form and sample were added in the csv file itself using the additional information given in the CDC website. The annotations that contributed to the additional attributes are given in table 2.

Phylum	Class	Form	Sample
Nematoda	Chromadorea	egg	intestine
Platyhelminthes	Cestoda	adult	feces
Apicomplexa	Enoplea	larva	muscle tissue
Amoebozoa	Conoidasida	trophozoites	bear meat
Sarcomastigophora	Rhabditophora	cyst	macro-organism
Ciliophora	Tubulinea	oocysts	brain tissue
	Trematoda	sarcocysts	wet mount
	Zooflagellate	gametocyte	eye
	Aconoidasida	schizont	bile duct
	Litostomatea	ookinete	na
		scolex	sputum
			blood
			liver
			lungs

Table 2: Attributes for the images.

2.3 Pre-processing the data

After annotating the data, it was observed that for some annotations, the images were very less (even 1 and 2 in some cases). As supervised machine learning is data intensive, these images were dropped. Only those images were kept which had at least 10 identical annotations in the dataset. The data was randomly split into 75% train and 25% test, with training and the test data containing independent and identical distribution of images (or the annotations). In total, the training data had 180 samples and the test data had 61 samples. The remaining organisms and their attributes are given in table 3 and 4 respectively.

Ascaris lumbricoides	Entamoeba histolytica	Plasmodium malariae
Cryptosporidium sp.	Enterobius vermicularis	Plasmodium ovale
Cyclospora cayetanensis	Giardia duodenalis	Plasmodium vivax
Dibothriocephalus latus	Hymenolepis nana	Sarcocystis sp.
Echinococcus granulosus	Plasmodium falciparum	Trichuris Trichuria

Table 3: Remaining species after pre-processing the data.

Phylum	Class	Form	Sample
Amoebozoa	Aconoidasida	adult	blood
Apicomplexa	Cestoda	cyst	feces
Nematoda	Chromadorea	egg	intestine
Platyhelminthes	Conoidasida	oocysts	liver
Sarcomastigophora	Enoplea	schizont	wet mount
	Tubulinea	trophozoites	
	Zooflagellate		

Table 4: Remaining attributes after pre-processing the data.

2.4 Prediction of species

For classification of parasites into their species, ResNet 50 (7), pretrained on ImageNet (8) data was used as the convolutional base. This was followed by a Global Average Pooling layer, a fully

connected dense layer having 512 units and output layer having 15 units (Fig 1), corresponding to the 15 species. In the first dense layer (with 512 units), Relu activation function was used, while in the output layer, softmax activation function was used. Dropout value of 0.25 was used in the fully connected layers for regularization. Adam's optimizer was used for updating the weights with a learning rate of 10^{-4}. The convolutional layers were frozen by looping over the layers and setting each freezing layers. The sample code for this in tensorflow is:

for layer in model.layers: layer.trainable = False

where model is ResNet 50, including only the convolutional layers and not the fully connected layers.

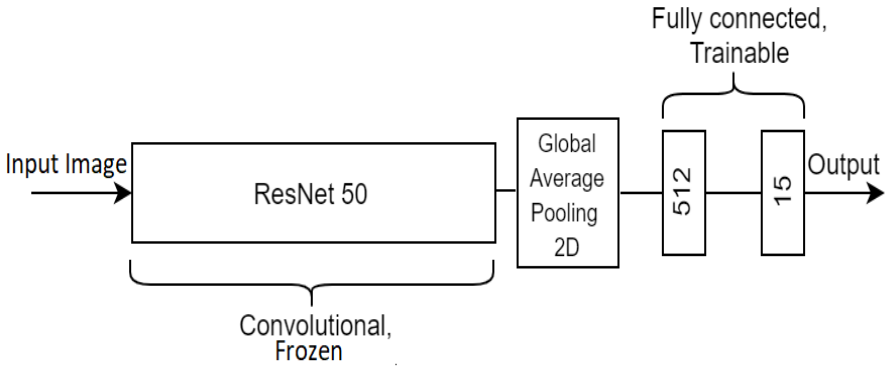


Figure 1: Architecture of the model used for species classification. The model has single input (image) and single output (species).

2.5 Prediction of attributes

The four attributes of the images have relationship among each other. For example, the organisms belonging to particular class will have particular forms, thereby will exist in particular environments from which the samples will be taken. Therefore, it becomes important to capture these relationships. Attention in neural network is one of the ways to capture the contextual relationship in sequence data, and are commonly used in sequence dependent tasks (9) (10). One of the popular ways to use incorporate such mechanisms is by using transformers models (11). Transformer models use multi-head self-attention (12) blocks to capture contextual relationships and have been effective for translation tasks. To use the transformer block in this study, the following modifications were made: (i) The positional embedding was removed to capture order free relationship. As using a predetermined order in the attributes can result in bias, it was necessary to remove the positional embeddings,

which accounts for the positional differences in the sequence. (ii) The fully connected feed forward connection was removed so as to keep minimum number of layers to prevent overfitting. The resulting structure formed the multi head self-attention block to account for order free relationships while making predictions. The neural network consisted of ResNet 50 as the base CNN architecture followed by four parallel fully connected layers, one for each attribute. To capture order-free relationship between the attributes, a multi-head self-attention module was used. For comparison, another model with the same architecture but without the attention module was used, also known as baseline. The model architecture is given in Fig 2. The convolutional layers were frozen as was done in section 2.4. Relu activation function was used, while in the output layer, softmax activation function was used. Dropout value of 0.25 was used in the fully connected layers for regularization. Adam's optimizer was used for updating the weights with a learning rate of 10^{-4}. The number of heads in the attention block was kept 8.

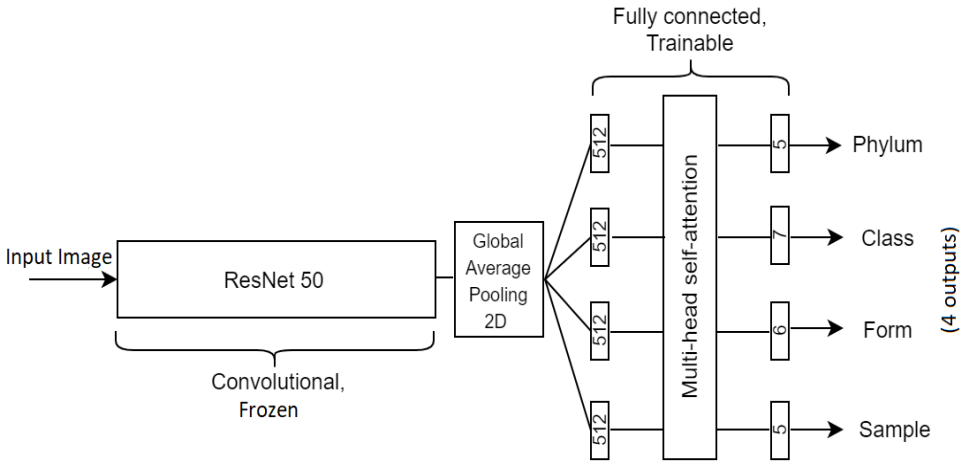


Figure 2: Architecture of the model used for attribute classification. The model has single input (image) and 4 outputs (attributes).

3 Results

The classification was performed on two stages: Predicting the organism and predicting its attributes. In the first stage, prediction of the organism was followed by applying Grad-CAM (13) to visualize the areas in the image that were utilized for making the predictions. Second stage of prediction was

performed by multi-label DNN incorporating a multi-head self-attention block. In this stage, the additional attributes of the images such as its phylum, class, form and sample were predicted. For predicting the organism, the neural network achieved an overall accuracy of 75.41%. The weighted-average F1 score was 0.73 and the macro-averaged F1 score was 0.76. The detailed results for all the organisms are reported in table 5 and the overall results are reported in Fig 3.

	Precision	**Recall**	**F1 score**
Ascaris lumbricoides	1.00	1.00	1.00
Cryptosporidium sp.	0.60	0.75	0.67
Cyclospora cayetanensis	0.67	0.50	0.57
Dibothriocephalus latus	1.00	1.00	1.00
Echinococcus granulosus	1.00	1.00	1.00
Entamoeba histolytica	0.67	0.67	0.67
Enterobius vermicularis	1.00	1.00	1.00
Giardia duodenalis	0.80	0.80	0.8
Hymenolepis nana	0.67	0.67	0.67
Plasmodium falciparum	0.67	1.00	0.80
Plasmodium malariae	0.75	0.86	0.80
Plasmodium ovale	0.00	0.00	0.00
Plasmodium vivax	0.67	0.50	0.57
Sarcocystis sp.	0.67	1.00	0.80
Trichuris Trichuria	1.00	1.00	1.00

Table 5: Detailed results for species classification (for all species).

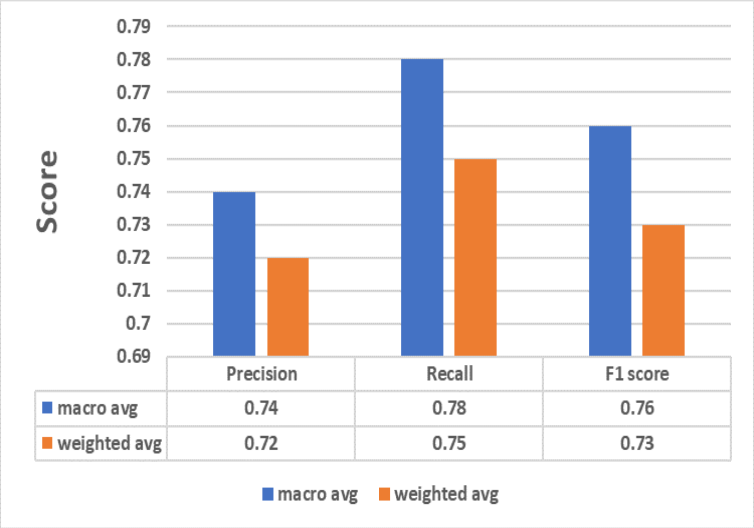


Figure 3: Overall results for species classification (for all species). The Score in the y-axis represents the respective averages of Precision, Recall and F1 Score in the plot. The macro avg is the unweighted average of respective scores (Precision, Recall or F1 Score) calculated for each species. In the weighted average, the weights are the sample numbers of individual species in test data and the average is calculated considering the weights.

For demonstration, the Grad-CAM visualization is shown for two predictions by the neural network (Fig 7,8). The heatmaps and the corresponding superimposed images show that the utilized parts for predictions involves the important part of the images (egg in case of Ascaris lumbricoides and infected cells in case of Plasmodium falciparum). For prediction of the attributes, the multi-label DNN involving the self-attention module performed better than the baseline model. While almost 66% of the times all the four attributes were predicted correctly with the baseline model, almost 74% of the times all the four attributes were predicted correctly with the model with self-attention (Fig 5,6).

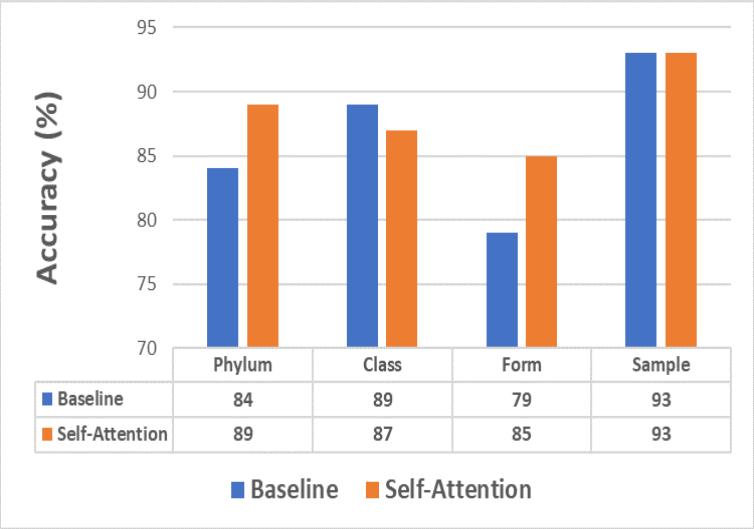


Figure 4: Attribute wise classification results by the attention based neural network and baseline (without attention).

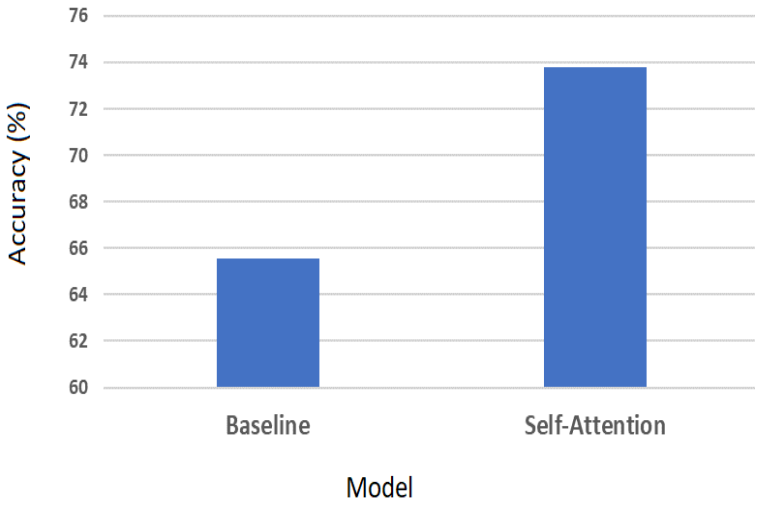


Figure 5: Performance Comparison for attribute classification. The graph denotes the % of times all the four attributes (Phylum, Class, Form and Sample) are predicted by the respective models. The Baseline denotes the model with the same architecture as Self-Attention but without the attention module.

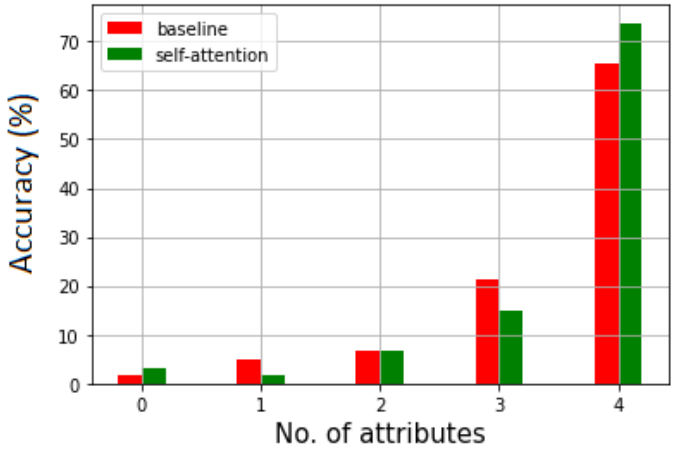


Figure 6: Accuracy in the y axis denotes % of times exactly n attributes predicted correctly (n=1,2,3,4). If the % correct prediction of exactly 4 attributes is higher, the corresponding % correct prediction of less than 4 attributes will be lower. Therefore, as the % correct prediction of all the 4 attributes for attention-based model is higher, the % correct predictions are lower for less than four attributes as compared to the baseline.

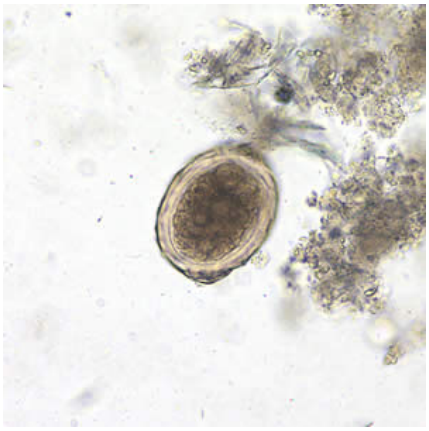
Phylum	precision	recall	f1-score
Amoebozoa	1.00	0.67	0.80
Apicomplexa	0.94	1.00	0.97
Nematoda	0.86	0.75	0.80
Platyhelminthes	0.73	1.00	0.84
Sarcomastigophora	0.67	0.40	0.50
Average Scores			
macro avg	0.84	0.76	0.78
weighted avg	0.89	0.89	0.88

Class	precision	recall	f1-score
Aconoidasida	1.00	1.00	1.00
Cestoda	0.73	1.00	0.84
Chromadorea	0.83	0.83	0.83
Conoidasida	0.88	1.00	0.93
Enoplea	1.00	0.50	0.67
Tubulinea	1.00	0.50	0.67
Zooflagellate	0.50	0.40	0.44
Average Scores			
macro avg	0.85	0.75	0.77
weighted avg	0.88	0.87	0.86

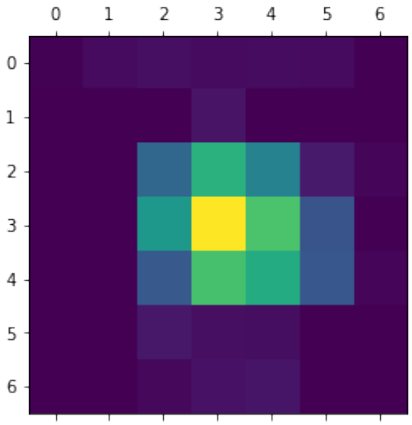
Form	precision	recall	f1-score
adult	1.00	1.00	1.00
cyst	0.80	0.73	0.76
egg	0.92	1.00	0.96
oocysts	0.93	1.00	0.97
schizont	0.00	0.00	0.00
trophozoites	0.77	0.85	0.81
Average Scores			
macro avg	0.74	0.76	0.75
weighted avg	0.81	0.85	0.83

Sample	precision	recall	f1-score
blood	1.00	1.00	1.00
feces	0.88	0.96	0.92
intestine	1.00	1.00	1.00
liver	1.00	1.00	1.00
wet mount	0.91	0.77	0.83
Average Scores			
macro avg	0.96	0.95	0.95
weighted avg	0.94	0.93	0.93

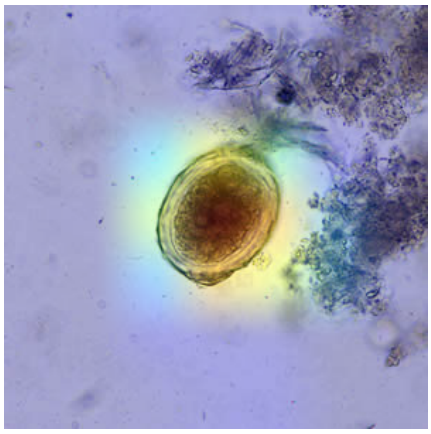
Table 6: Attribute Classification Results for attention based neural network. The Average Scores are the macro and weighted averages for Precison, Recall and F1 Scores (under the respective column headings). The macro avg is the unweighted average of respective scores (Precision, Recall or F1 Score) calculated for each class of the particular attribute. In the weighted average, the weights are the sample numbers of individual class of the attribute in test data and the average is calculated considering the weights.



(a) Original Image



(b) Heatmap



(c) Superimposed image

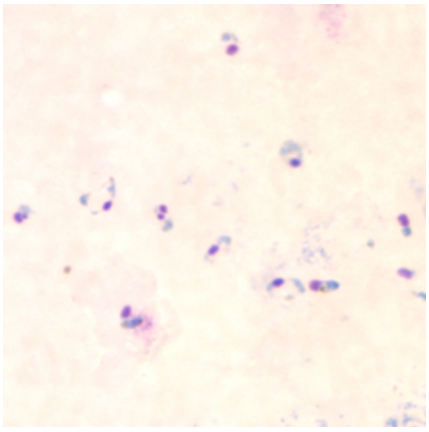
Figure 7: Grad-CAM visualization for the organism Ascaris lumbricoides. (a) : Original image of Ascaris lumbricoides as downloaded from the CDC website. (b): The heatmap created with Grad-CAM showing the important parts of the image. (c): The image obtained after superimposing the heatmap (b) on the original image (a). The prediction for the orginal image (a) was as follows:

Species: Ascaris lumbricoides

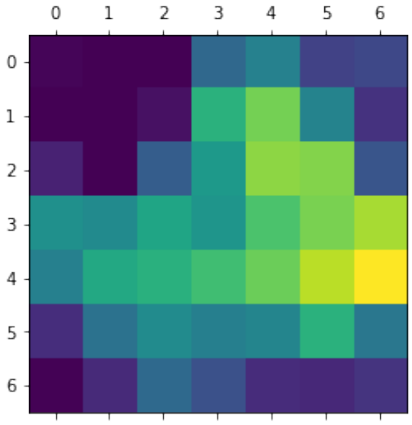
Attributes: Phylum: Nematoda, Class: Chromadorea, Form: egg, Sample: feces

(The species was predicted by neural network as mentioned in section 2.4 and the attributes were predicted by attention based neural network.)

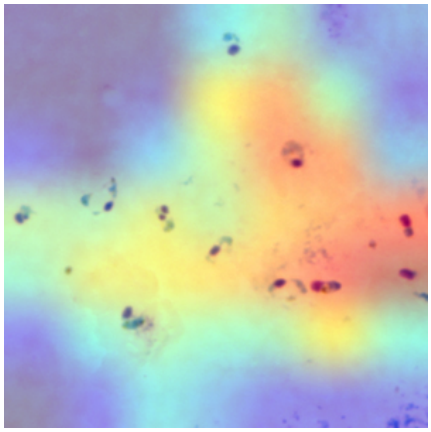
The examples for two predictions are shown in the Fig 7 and Fig 8. In the first stage, the type of species and the corresponding important regions in the image are analysed, and in the second level the rest of the attributes corresponding to the image are analysed.



(a) Original Image



(b) Heatmap



(c) Superimposed image

Figure 8: Grad-CAM visualization for the organism Plasmodium falciparum. (a) : Original image of Plasmodium falciparum as downloaded from the CDC website. (b): The heatmap created with Grad-CAM showing the important parts of the image. (c): The image obtained after superimposing the heatmap (b) on the original image (a). The prediction by the neural network for the orginal image (a) was as follows:

Species: Plasmodium falciparum

Attributes: Phylum: Apicomplexa, Class: Aconoidasida, Form: trophozoites, Sample: blood

(The species was predicted by neural network as mentioned in section 2.4 and the attributes were predicted by attention based neural network.)

The data and the code are availabe at:

https://github.com/Devanshpandey/Learn-parasitology-and-diagnose-with-Deep-Learning/

4 Discussion

In this study, parasitology image data was downloaded and annotated. Then two stage prediction was performed, first stage predicting the species with visual explanations and the second stage predicts the additional attributes of the images. For predicting the attributes, a DNN with multi-head self attention model was used and was found to perform better than the baseline. The results despite the low sample numbers show the potential application of deep learning approaches in this direction of study.

5 References

1. Deep learning architectures for multi-label classification of intelligent health risk prediction. al., Maxwell et. s.l. : BMC Bioinformatics, 2017.

2. A Novel Deep Neural Network Model for Multi-Label Chronic Disease Prediction. al., Zhang et. s.l. : Frontiers in Genetics, 2019.

3. Fully DNN-based Multi-label regression for audio tagging. al., Yong et. s.l. : arXiv, 2016.

4. Multi-label learning based deep transfer neural network for facial attribute classification. al., Zhuang et. s.l. : Pattern Recognition, 2018.

5. Multi-label clinical document classification: Impact of label-density. al., Blanco et. s.l. : Expert Systems with Applications, 2019.

6. https://www.cdc.gov/dpdx/az.html.

7. Deep Residual Learning for Image Recognition. al., He et. s.l. : arXiv, 2015.

8. ImageNet: A large-scale hierarchical image database. al., J. Deng et. 2009, 2009 IEEE Conference on Computer Vision and Pattern Recognition.

9. Attention in Natural Language Processing. Andrea Galassi, Marco Lippi,Paolo Torroni. s.l. : arXiv, 2019.

10. Coarse-to-Fine Attention Models for Document Summarization. Jeffrey Ling, Alexander Rush. s.l. : Proceedings of the Workshop on New Frontiers in Summarization, 2017.

11. Attention Is All You Need. al., Vaswanai et. s.l. : arXiv, 2017.

12. Analyzing Multi-Head Self-Attention: Specialized Heads Do the Heavy Lifting, the Rest Can Be Pruned. al., Elena et. s.l. : arXiv, 2019.

13. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. al., Selvaraju et. s.l. : International Journal of Computer Vision, 2019.