USING CONVOLUTIONAL NEURAL NETWORK MODELS TO IDENTIFY PARASITES

KEREM DOGAN

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE

DEPARTMENT OF
COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

STUDENT NUMBER

838468

COMMITTEE

dr. Merel Jung
dr. Sharon Ong

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

June 24, 2022

ACKNOWLEDGMENTS

USING CONVOLUTIONAL NEURAL NETWORK MODELS TO IDENTIFY PARASITES

KEREM DOGAN

Abstract

The goal of this project is to use multiple pre-trained convolutional neural networks(CNN) on the dataset named "Microscopic Images of Parasites Species" by Li et al, (2020) to identify parasites and compare the performances of each model. The main question that the paper is trying to answer is whether can CNN is used to classify different parasites. In this research, CNNs are used to classify the eight different parasites that are in the data set. Then the models are evaluated on their performances to be compared. The dataset contains 34.298 images of 8 different parasites and some indifferent magnification. Due to the data set being unbalanced 1400 random images were chosen to balance out the data. The results show that the convolutional neural networks can accurately identify parasites. However, from time to time the performances differ from one another. And the model should be chosen according to the preferences. Therefore, comparing the convolutional neural networks can give insight into efficiency and how the models work.

1 INTRODUCTION

With the recent pandemic, the topic of health and disease becomes more frequent in the daily lives of many. According to CDC (2022), parasites are organisms that live on or in another creature and feed on or at the expense of that organism. Parasites come in a variety of forms. However, three types of parasites are particularly hazardous to humans and they are Ectoparasites, helminths, and Protozoa. Ectoparasites are arthropods that infect hosts by attaching themselves. Helminths are a worm that enters the host's body and then grows in the host's body. Protozoa are microscopic parasites that can reproduce inside the host's body and can be transmitted by bugs. Parasitic diseases annually cause a lot of demise and despair (Momčilović et al., 2019). An example of these parasitic diseases is malaria.

"Malaria affected an estimated 219 million people causing 435,000 deaths in 2017 globally"(Talapko, Škrlec, Alebic´, Jukic´, Vc˘ev, 2019, pp.1).

The Dataset "Microscopic Images of Parasites Species" by Li et al, (2020) consists of images of eight different parasites. The parasites in the dataset are Babesia, Leishmania, Leukocyte, Plasmodium, RBCs(parasites that affect red blood cells), Toxoplasma, Trichomonad, and Trypanosome. First of all, Babesia also known as Babesiosis is a kind of RBC(red blood cells) parasite. Where the parasite infects the host's red blood cells. The diagnosis of this of babesia is usually achieved by examining the blood cells under a microscope (CDC 2018). Also according to CDC(2018), babesiosis can be dangerous and even fatal if the host has a weak immune system. The leishmania is also known as leishmaniasis. The Leishmania parasite is transmitted by sand flies. According to CDC(2020), leishmaniasis has two different variants which are cutaneous leishmaniasis and visceral leishmaniasis. Cutaneous leishmaniasis can cause skin sores that aren't that harmful but, visceral leishmaniasis can affect internal organs which could cause severe harm to the host's organs. Leukocyte is the white blood cells which is a part of the immune system of the body. Leukocyte parasites such as can infected the white blood cell. According to CDC(2022), Plasmodium has 156 different species and the most known case of Plasmodium is Malaria. Specifically, the malaria parasite is carried and transmitted by mosquitos.RBCs (red blood cells) are parasites that infect the host's red blood cells as mentioned above Babesia is a prime example of RBCs. Toxoplasma is also known as Toxoplasmosis is a parasite that is mostly transferred from eating under-prepared/cooked food. According to CDC(2018), Toxoplasmosis is one of the major causes of death in terms of foodborne diseases. Trichomonad also known as Trichomoniasis is a sexually transmitted parasite (CDC,2022). According to CDC(2022), in the USA there are over 2 million cases alone in 2018. Trypanosome, also known as Trypanosomiasis and "sleeping sickness". Trypanosome is transmitted by flies which is local to Africa(CDC,2022). According to CDC(2022), if trichomoniasis is not treated it might become deadly.

As seen in the previous paragraph these parasites are dangerous and even could cause a lot of harm and fatalities. To overcome these parasitic diseases, early diagnosis is one of the crucial points. However, in underdeveloped areas, the early diagnosis becomes problematic due to a lack of experts and assets. (Momčilović et al., 2019). To get ahead of these parasitic disease outbreaks there must be an accessible and accurate kind of diagnosis method. According to Momčilović et al(2019, pp.291), microscopes are one of the cheapest methods to detect parasites. However, this method requires an expert who prior experience and parasitological knowledge. However, on paper (Saeed, Jabbar, 2017, pp.7) usage of mobile phone were

suggested to detect parasites. This idea made the process accessible so that a microscope wasn't required and everyone who has an access to a mobile phone could use this method. Yet still, it required a manual operation which caused a lot of false results. The lack of professional operators and precision could be solved with the help of deep learning. According to Jadhav and Yadav (2019, pp.4), a convolutional neural network(CNN) can outperform two human experts. The convolutional neural network(CNN) is a neural network that is good for analyzing images. CNN is useful and efficient since the network by itself will observe and learn from the features of the images (Jadhav, Yadav, 2019). Also, CNNs can detect features more accurately in a shorter time which makes the whole process more efficient (Jadhav, Yadav, 2019).

1.1 *Research Question*

As stated in the introduction, the study will look into classification of different protozoa parasites with the usage of multiple pretrained convolutional neural networks. Which entails the research question of:

RQ1 *To what extent can parasites be classified from microscopy images using pretrained convolutional neural networks?*

As a follow up to this question, since there are multiple pretrained CNNs their accuracy and their methodology will be different from one another. Therefore, some of them would yield better accuracy and efficiency. Which entails the following sub-question:

RQ1.1 *which pretrained convolutional neural network should be used?*

In the end, the results showed that the pretrained models can accurately predict and classify the right parasites. However, the pretrained models even without fine tuning showed a high f1 score and accuracy values. Which can mean either the models are to complex for the dataset or the dataset is simple so it is easy for the models to classify the parasites. Yet, there are still small but notable differences between the pretrained CNN models.

2 RELATED WORK

The early detection of these parasitic diseases is important to stop an outbreak from occurring. There are multiple methods to detect these types of parasites. The paper by Talapko, Škrlec, Alebić, Jukić, Včev (2019, pp.2, pp.3) brings up different methods of detecting malaria parasites and the methods were split into two categories such as direct and indirect methods.

Direct methods look for the existence of parasites, and indirect methods check antibodies to determine the presence of the parasite. The methods are microscopic analysis, molecular tests, rapid diagnostic tests, indirect immunofluorescence, and ELISA. The rapid diagnostic test, microscopic analysis, and molecular tests use a direct method. The advantages of the rapid diagnostic test are that it is fast and easy yet the disadvantages are that it is inaccurate and less sensitive. The microscopic analysis is fast and cheap but it requires equipment and a professional operator. Finally, the molecular tests are sensitive and accurate but it is expensive and it takes a long time. On the contrary, indirect immunofluorescence and ELISA use indirect methods. Both methods are accurate and definitive. Despite this, both ELISA and indirect immunofluorescence take a long time to diagnose, require professional operators, and requires advanced tools. Therefore, in the underdeveloped parts of the world, it would be more beneficial to use a direct method that is both easy to access, cheap, and easily operatable. According to Momčilović et al, (2019, pp.291), one of the cheap ways of detecting these parasites is using a microscope. However, also Momčilović et al (2019, pp.291) mention that detecting these parasites requires the operator to have parasitological knowledge and experience with the parasites. But, the operators can be replaced by developing deep learning algorithms that are as capable as the parasitological operators. (Das et al, 2013) Therefore, the microscopic analysis would become an efficient method to detect parasites even in underdeveloped parts of the globe. One of the innovative ways of diagnosing parasites was by using phones(Saeed, Jabbar, 2017, pp.7). The advantage of using phones for diagnosis is that it is easily accessible. During the research, there were different methods were used. These methods were standalone smartphone technology, smartphone-assisted microscopy, and attachable lens microscopy. The standalone smartphone technology used the phone's camera and downloadable applications(apps) for diagnosis. The problem with this method was that it wasn't sensitive enough. Secondly, smartphone-assisted microscopy consisted of a handheld microscope and a smartphone. Finally, attachable lens microscopy uses a camera lens that acts as a microscope. The difference between attachable lens microscopy and smartphone-assisted microscopy is that smartphone-assisted microscopy uses a stand-alone microscope and attachable lens microscopy uses a lens that acts as a lens. Overall, using phones for diagnosis can be effective and efficient. Additionally, this would solve the problem of equipment requirement of the microscopic analysis that was mentioned since most people have access to a smartphone. However, these methods require manual sorting which sometimes caused false results. Also, there were some problematic factors like hygiene, limited field of view and manually operating the slides. A similar idea to Saeed,

Jabbar's(2017) research, Fuhad et al(2020) aimed to develop an application that uses a deep learning system to automatically detect malaria parasites from blood smears. An interesting point to the methodology is that rather than using only deep learning the experiment involved the use of both deep learning and machine learning. They did it by using CNN models with either KNN or SVM. This way they were able to achieve high precision values. Additionally, Fuhad et al(2020) develop a mobile application and a web-based application to use their model. And, both of the models were able to accurately detect malaria parasites. Usage of mobile phones and web-based applications could probably solve accessibility problems of methods that are mentioned above. In today's world, almost everyone has a smartphone or has access to the internet.

Another way of parasite diagnosis is Digital PCR(dPCR) by Pomarti et al,(2019). The main idea behind the dPCR is achieved by using Poisson statistics, amplifying a single DNA template from maximum diluted samples, resulting in amplicons that are solely generated from one template (Pomarti et al,2019, pp.1511). Additionally, dPCR generates digital signals that are linear therefore it can allow the statistical analysis of the PCRs. Also, it is mentioned that while detecting malaria parasites a lot of systems use tools with low sensitivity since malaria causes high blood parasitic loads which are easy to detect. However, these tools with low sensitivity aren't enough to detect more developed parasites such as 'chronic malaria' in that case, it suggested that the dPCR with high sensitivity could be a useful tool. dPCR showed a better performance since it can detect the high sensitivity of PCR can detect the low parasitism of Schistosomiasis. To conclude, dPCR showed a higher performance than qPCR on tasks where higher sensitivity is required. However, dPCR has a lot of complications and limitations to overcome for its implementation.

The paper by Ricciardi et al(2014) mentions the usage of serology while detecting parasites. According to CDC(2022), serology is used to detect antibodies or parasite antigens. In the paper, it is mentioned that serology improved the diagnosis of parasites. Additionally, the serology on leishmania discovered a lot of antigens that could be used on detecting the leishmania parasite. The best anti-body rapid test was able to detect Leishmania parasites in 15 minutes. However, the antibody only showed this performance in Asia and not in Africa (Ricciardi et al,2014). Also, it is said that while diagnosing toxoplasmosis serology plays a huge role since most cases of toxoplasmosis are asymptomatic. Also, it is mentioned that while detecting trypanosomiasis in West Africa the anti-bodies that were used to diagnose gave false-positive results. Although, serology has its benefits results are checked by microscopy to confirm the result. It is also mentioned that microscopy is still the gold standard for diagnosis.

However, serology is still used and still, several kits rely on serology for diagnosis.

In the research by Das et al(2013), machine learning algorithms were used for screening malaria parasites. The algorithms that were used were Naïve Bayes and SVM. In the end, the machine learning algorithm that was used can successfully classify cells that are infected by malaria. Finally, the researchers concluded that machine learning plays a big role in quantitative microscopy and that it may be used in telemedicine to give a rapid diagnosis in remote locations where pathologists are few (Das et al, 2013, pp.105, pp.106). Similar to the previous paper that was mentioned Umer et al(2020) also worked on the malaria parasite yet the difference was that rather than machine learning a stacked convolutional neural network was used. Throughout to process, it is seen that preprocessed CNNs yielded a higher accuracy and f1 value. Additionally, as the stack size increases a higher accuracy and f1 value were observed. Finally, the paper concluded that this research shows that CNN features outperform hand-crafted features (Umer et al,2020, pp.93790).

As seen from the mentioned research above there are examples of deep learning and machine learning used to detect parasites. And, the usage of deep learning and machine learning for parasite detection is suggested. Yet, the research is based on malaria parasites and is only limited to malaria parasites. Therefore, using the dataset by Li et al, (2020) a more complete system that can efficiently recognize and differentiate similar parasites could be developed. Additionally, this method of diagnosis and classification can be beneficial to use since it is accessible and efficient. Since the task requires good feature detection a CNN model would be used. Additionally, the system must be accurate with the outcome since there shouldn't be any false results. To overcome this problem a well-trained model is required therefore a pre-trained model would be used. Benbihi et al,(2019) concluded that the pre-trained outperforms state-of-the-art detectors, it is also easy to adapt it to pre-existing data and improve the performance. Also, since the goal is to achieve the highest possible accuracy and f1 value we should run multiple pre-trained models.

3 METHOD

3.1 *Brief Overview*

The main goal of the research is to identify and classify different parasites. To overcome this goal a system with good feature detection is required hence a convolutional neural network is used. Additionally, since the goal requires as minimum false results as possible multiple pre-trained models

were used. Pre-trained models are proven to improve performance and since the models use different algorithms multiple of them were used to determine the best performing model. These pre-trained models were used from the fastai library since fastai utilizes transfer learning. Kornblith et al,(2018) suggested that convolutional neural networks that use training data supplied through transfer learning might significantly surpass these hand-engineered features.

The dataset by Li et al, (2020) includes eight types of different parasites and they are Babesia, Leishmania, Leukocyte, Plasmodium, and RBCs(parasites that affect red blood cells), Toxoplasma, Trichomonad, and Trypanosome. There is a total of 34.298 images on the dataset and the dataset itself isn't balanced. Both leukocytes and Toxoplasma have images on two different magnification levels. But Toxoplasma at 400x zoom and Leukocyte at 1000x zoom was removed from the dataset since Toxoplasma at 400x zoom had a poor resolution and Leukocyte at 1000x zoom had a small sample size. Since the data set was imbalanced 2172 pictures of parasites were randomly picked to be used. Also, it effectively helped with computational power and efficiency Then the data was split into 60/20/20 to training/validation and test sets. The data is optimized using learning rate in training/validation set in order to develop a better performing model. The final evaluation of the models are is done on the test sets. During the optimization only learning rate hyper parameter is used.

3.2 *Software*

For this experiment, the Python software(Van Rossum and Drake, 2009) will be used in order to build the script that's needed. The Python software is acessed by Jupyter notebooks Via Anaconda Navigator. The scripts are used to acess and edit the dataset, development of training/validation sets, running pre-trained models and visualize the findings. The editing of the dataset is mostly done by NumPy.(Harris,, Millman, van der Walt, et al, 2020) The visualisations of the findings were done by Matplotlib(Hunter,2007). Additionally, the CNN models and the pre-trained models were used from the fastai library(howard,2018). Finally OpenCV (Bradski, G. 2000) ,Pathlib, Fastbook(Howard et al, 2020), Os ,string and glob libaries were used.

3.3 *CNN*

The convolutional neural network (CNN) is an artificial neural network that is useful for analyzing images. One of the major reasons for convolutional neural networks being useful is that they will detect and learn the

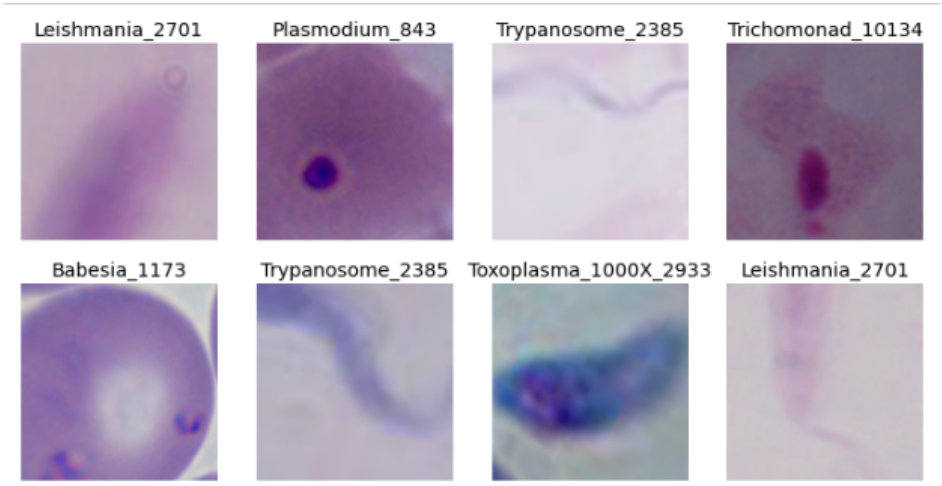

Figure 1: Example from the dataset (Li, Sen; Zhang, Yang, 2020)

attributes of the visuals by themselves(Jadhav, Yadav, 2019). Additionally, convolutional neural networks can extract features efficiently which makes the process more effective and easier (Jadhav, Yadav, 2019). The convolutional neural network has two major parts which are classifiers and feature extractors. In the feature extractor layer, the output from the previous layer is used as an input and the new output is given as an input to the next layer. The convolutional neural network has three different layers and they are convolution, max-pooling, and classification layers (Zahangir Alom et al, 2018). The convolution and max-pooling layers are mostly found at the beginning of the model. In the convolution layer, the output of the previous layers is transformed into a learnable kernel. The max-pooling layer downsamples the incoming input however, only the dimensions of the input are downsampled therefore the input and the output have an equal size. The output of these convolution and max-pooling layers are known as feature mapping and they are acquired mostly from the outputs of the previous layers. More feature mapping means a better understanding of the features therefore, the model gets a higher accuracy. (Zahangir Alom et al, 2018) The final layer is the classification layer and is a network that is fully connected. The input for the classification layer is gathered from the last layer of the model to get the feature map of the final layer. Some models often have multiple classification layers to make it easier for computation. (Zahangir Alom et al, 2018)

3.4 *Using pre-trained models on the dataset*

In a similar case, Umer et al(2020) used stacked CNNs to identify malaria parasites. The CNN used in this research is pre-trained since higher f1 and accuracy scores are aimed. To use the pre-trained model on the dataset, the pre-trained model goes through transfer learning and starts learning the given "Microscopic Images of Parasites Species" dataset. Then, the weights of the pre-trained model are passed on to a new model then the new model begins to train with the given dataset(Bens Pardamean et al, pp.402). The dataset is trained with the fastai's "fine tune" function. The function reconditions the weights of the following layers quicker than the previous layers. The data is then fine-tuned throughout a single epoch in order to Make sure everything is set correctly. Later, it is used to choose a proper learning rate for the model. In the models non of the layers were re trained in any layers only the last layer was trained. Finally, in this research 8 pre-trained models were used and they are Alexnet, Resnet (34,50,101), Densenet(169,201), VGG16, and squeezenet1.1. While the interpretation of the individual confusion matrix of the models, the minimum confusion was set to 3. So if the CNN mixed a parasite 3 times then it is assumed that the CNN confuses them. This is done to avoid false results. Additionally, if the model confused the parasites then it would go under a fine-tune to get avoid confusion and to get a higher f1 and accuracy value.

3.5 *AlexNet*

AlexNet is a CNN model made by Alex Krizhevesky in 2012. It is said to be one of the biggest development for image classification and recognition (Zahangir Alom et al, 2018). The Alexnet model has 7 layers in total. But, it only has 3 convolution layers and 2 fully connected layers(2 classification layer) (Zahangir Alom et al, 2018).

3.6 *Resnet*

Resnet is created by Kaiming He and the goal of the project is to create a model which didn't got effected by vanishing gradient problem. The Resnet model has multiple variations which have different values (Zahangir Alom et al, 2018). The ones that are used are Resnet 34, Resnet 50 and Resnet 101. The models only have 1 fully connected layer and the rest of the layers are convolutions layers. In the case for Resnet 50, there are only 1 connected layer and 49 convolutions layers (Zahangir Alom et al, 2018).

3.7 *Densenet*

Densenet is model that is developed in 2017 and it is based around CNN layers that are densely connected to one another(Zahangir Alom et al, 2018).The main idea behind Densenet is that the output of layers are connected to all of the following layers. Then the combined layers become a dense block. In a densenet model there are multiple dense blocks and the blocks are connected to one other in order to navigate trough the dense blocks. An advantage of the Densenet is that it can reuse features, which decreases the network pattern.(Zahangir Alom et al, 2018) The Densnet similar to Resnet have multiple variants such as Densnet 121, 161, 169, 201. However, in the research, only Densenet 169 and 201 are used.

3.8 *VGGNet*

VGGnet also known as Visual Geometry Group is a important CNN model. Since the VGG model has proven that depth of a network could help with achieving higher accuracy. The VGG model also have multiple variants such as VGG 11,16,19. In this research VGG 16 is used and it is consisted of 13 convolution layers and 2 fully connected layers(2 classification layer).(Zahangir Alom et al, 2018)

3.9 *SqueezeNet*

SqueezeNet is model that is mainly built for machines that lack the computational power. Mainly SqueezeNet is focused to be used on mobile devices. Therefore, the main goal of SqueezeNet is to reduce the size of the network. The SqueezeNet has to versions and they are SqueezeNet 1.0 and 1.1. The major differences between them are that 1.1 version has less strides and less filters which makes the network smaller. In the research SqueezeNet 1.1 is used (Koonce, 2021).

4 RESULTS

Table 1: Best Performances of Pre-trained Models

Models	train loss	valid loss	error rate	accuracy	f1 score	roc auc score	time
Alexnet	0.174	0.094	0.013	0.986	0.986	0.998	00:24
Resnet 34	0.226	0.088	0.031	0.968	0.9689	0,999	02:21
Resnet 50	0.130	0.042	0.020	0.979	0.979	0.999	04:47
Resnet 101	0.133	0.051	0.017	0.982	0.982	0,999	08:07
Densenet 169	0.122	0.062	0.024	0.975	0.975	0.999	07:14
Densenet 201	0.045	0.017	0.006	0.993	0.993	0.999	49:44
VGG16	0.277	0.085	0.031	0.968	0.9689	0,999	08:23
Squeezenet 1.1	0.188	0.037	0.013	0.986	0.986	0.999	00:49

4.1 *AlexNet*

In the training set, as seen on the figure 2 the initial findings AlexNet scored a good accuracy and f1 value of 0.965. Since, the train loss has a significantly bigger value than valid loss which means the model is under fitting. Initially both train loss and error rate has a high value and can cause problems yet, in the end it seems that both of the values decreased drastically which means the model was successfully at predicting. Additionally, there is a high roc auc score which means AlexNet was good at distinguishing positive class and the negative class. The confusion matrix shown in figure 3 shows that AlexNet confused Babesia and RBC's. Therefore, to get over this confusion a better learning rate was picked. As seen in figure 4 the suggested was 1.10e-04 and when the model is run with the suggested learning rate as seen at figure 6 AlexNet confused Babesia and RBC's only once which is an improvement. Additionally, the accuracy and the f1 values are at 0.986 which shows that are model is improved. However, there are still a Under fitting problem since in the figure 5 the training loss is significant bigger than the valid loss. To summurize, the AlexNet model is a good at predicting the given parasites but there is a under fitting problem. In the test set Alexnet predicted 329 out of 361 parasite images correctly with an accuracy score of 0.9113.

Alexnet confused the Trichomonad parasite with the Leishmania parasite the most.

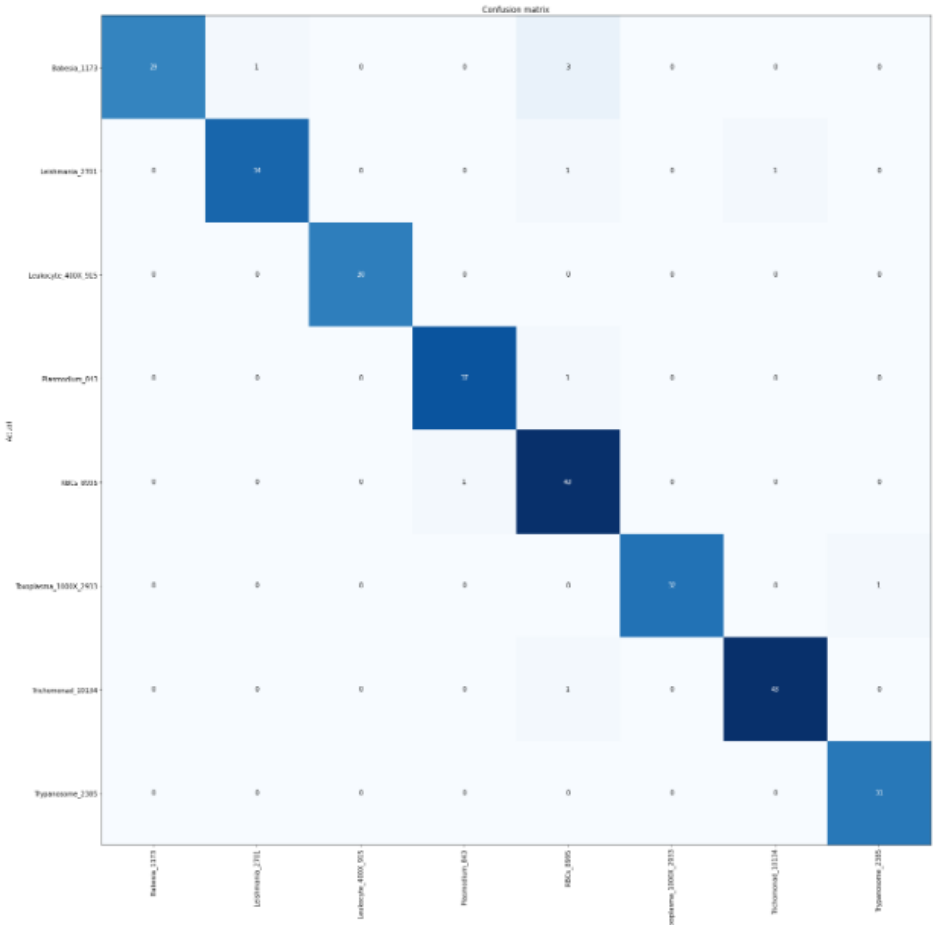

Figure 2: Confusion matrix of Alexnet

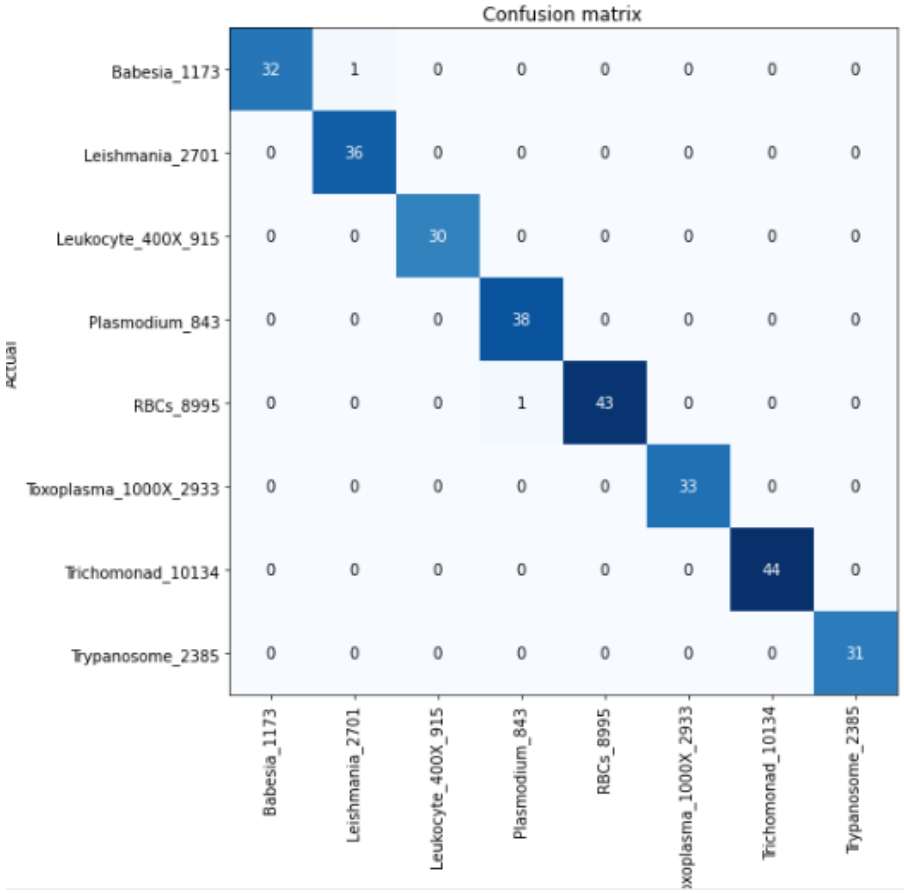

Figure 3: Confusion matrix of Alexnet with the best learning rate

4.2 *Resnet*

In this research three different resnet models were used and they are resnet 34,50 and 101. The models were used both in training and test set. In the training set, the figures 7,9,11 it is seen that as layers of the resnet increases both accuracy and f1 values increases. Therefore, resnet 101 has the best accuracy and f1 score of 0.982 and resnet 34 has the worst accuracy and f1 score of 0.968 which is still really good. Relatively all of the models have significantly high roc auc score which means that the model can detect positive and negative classes with high accuracy. When looked at the figures 8,10,12 there are some confused parasites yet the amount that is confused isn't significant. Therefore, there is no need to check for the best learning rate. All of the models have a significantly higher train loss value than valid loss which means all of the models are under fitting. Additionally, resnet 34 has high train loss and valid loss values which shows the model performed poorly in the train and validation sets. To conclude all of the models performed really well and the best performing one was resnet 101. Also, On all of the models some under fitting can be observed. In the test set resnet 34 predicted 301 out of 361 parasite images correctly with an accuracy score of 0.833. Resnet 34 confused the RBC parasite with the Babesia parasite the most. the resnet 50 predicted 293 out of 361 parasite images correctly with an accuracy score of 0.8116. Resnet 50 confused the Plasmodium parasite with the RBC parasite the most. Finally, the resnet 101 model predicted 283 out of 361 parasite images correctly with an accuracy score of 0.7839. Resnet 101 confused the Trypanosome parasite with the RBC parasite the most.

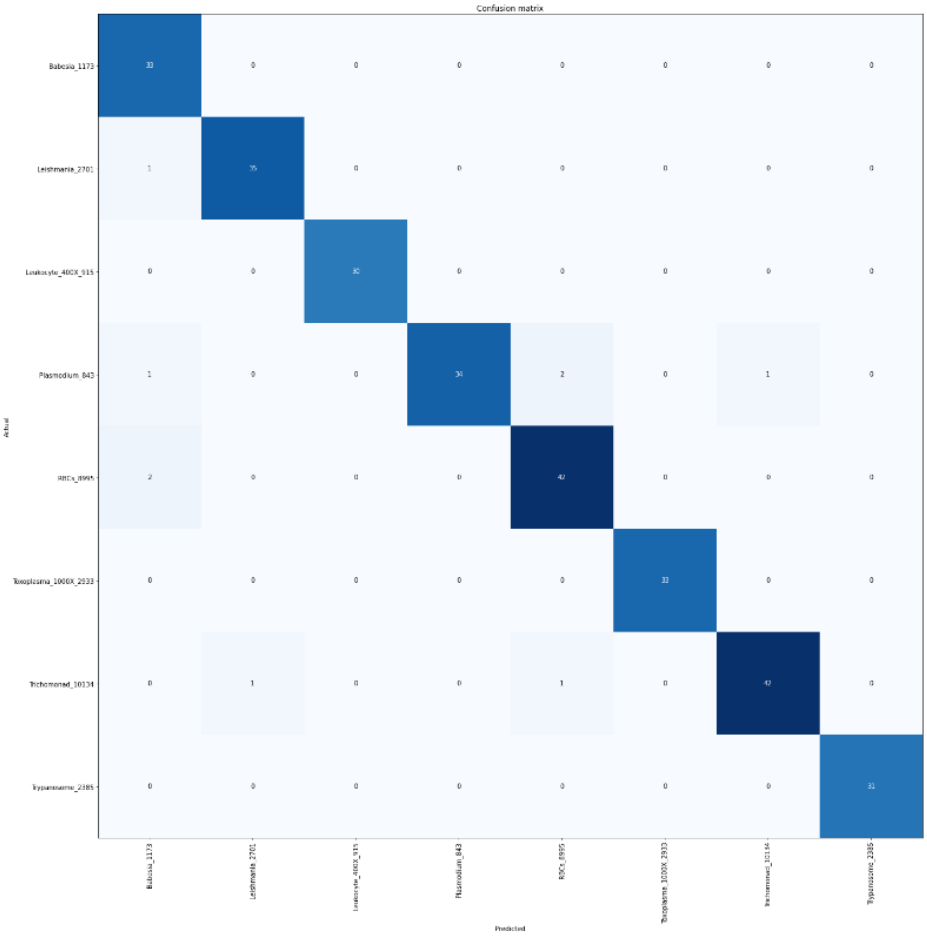

Figure 4: Confusion matrix of ResNet34

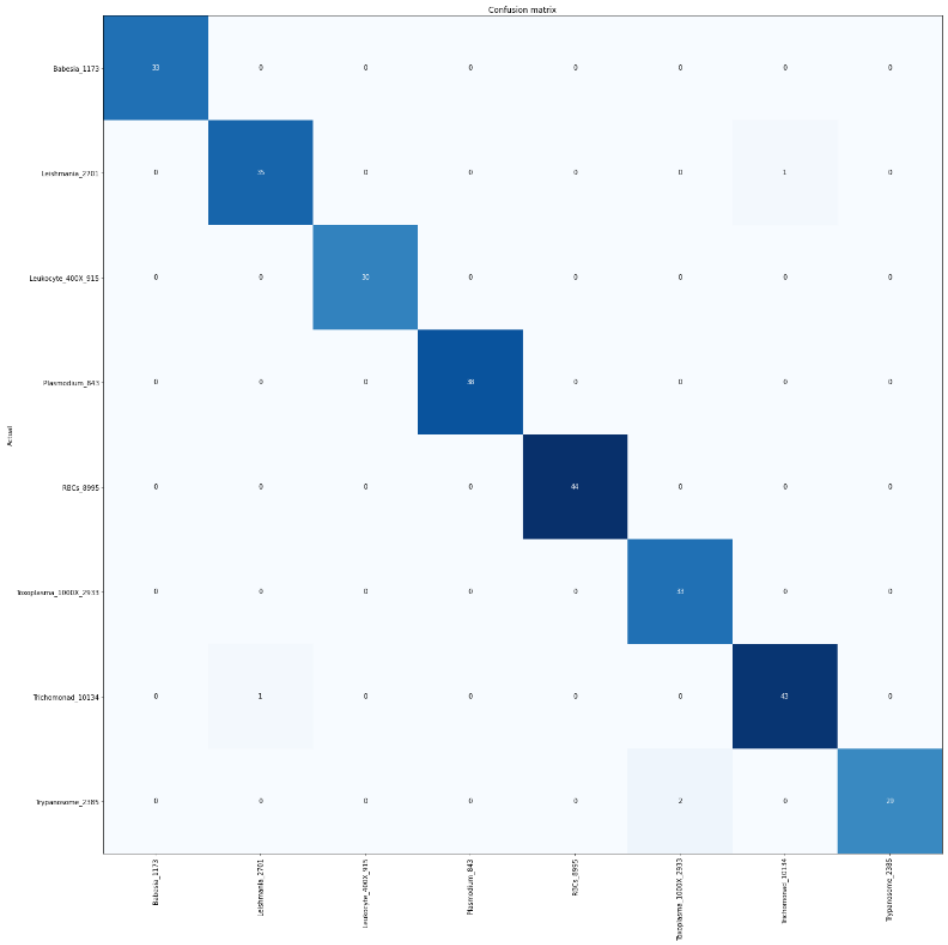

Figure 5: Confusion matrix of ResNet50

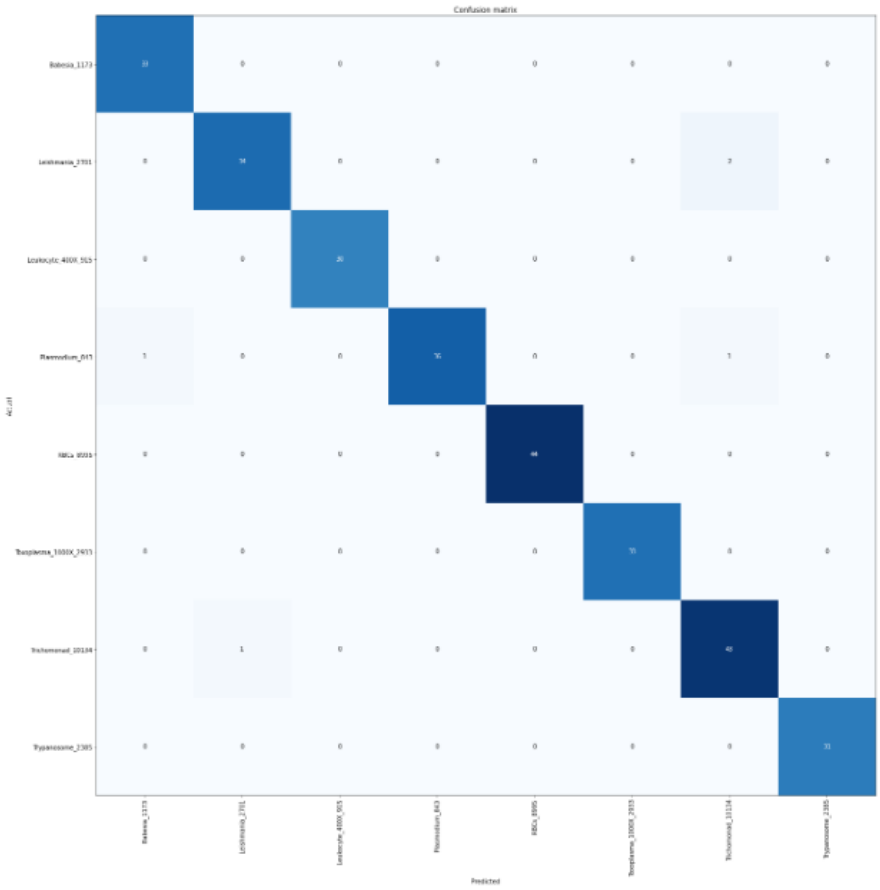

Figure 6: Confusion matrix of ResNet101

4.3 *Densenet*

There are two different Densenet model that was used and they are densenet 169,201.In the training set, the Figure 13 shows that densenet 169 has a high accuracy and f1 value of 0.975. The model has a really high roc auc score of 0.999 which means it classified positives and negatives almost flawlessly. Additionally, train loss is significantly bigger than the valid loss which entails under fitting. Also like densenet 169 densenet201 performed really well.According to the figure 15 The densenet 201 had a high accuracy score of 0.961 and a f1 score of 0.962. Also the model had a similar roc auc score of 0.999 like densenet 169. However the densenet 201 have higher error rate. However, in the figure 16 it is seen that densenet 201 confused Rbcs with Babesia and leishmania with babesia. Therefore, a better learning rate was picked.On figure 18 with the new learning rate densenet 201 had a really high accuracy and f1 score of 0.993 while maintaining the roc auc score of 0.999. Also, the error rate drastically decreased from 0.03 to 0.006. Both the train loss and valid loss have a small value which shows that densenet201 performed well both on validation and training set. Even though, the values of train loss and valid loss are low still train loss is bigger than the valid loss which shows under fitting. To summrize, both densenet 169,201 performed really well. However, both models encountered under fitting problem. Additionally, running densenet models takes a lot of time to process as seen on figure 13,15 and 18. In the test set densenet 169 predicted 353 out of 361 parasite images correctly with an accuracy score of 0.9778. densenet 169 confused the Toxoplasma parasite with the Trypanosome parasite the most. And, the densenet 201 predicted 326 out of 361 parasite images correctly with an accuracy score of 0.9430. densenet 201 confused the Toxoplasma parasite with the Trypanosome parasite the most.

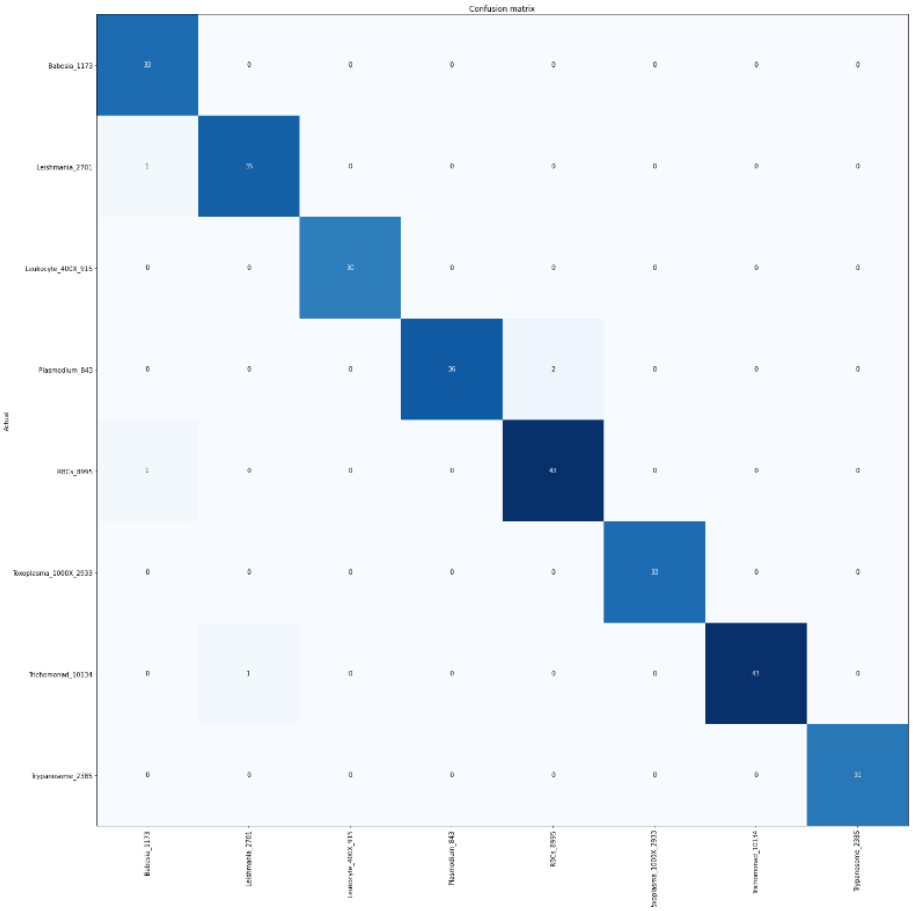

Figure 7: Confusion matrix of densenet 169

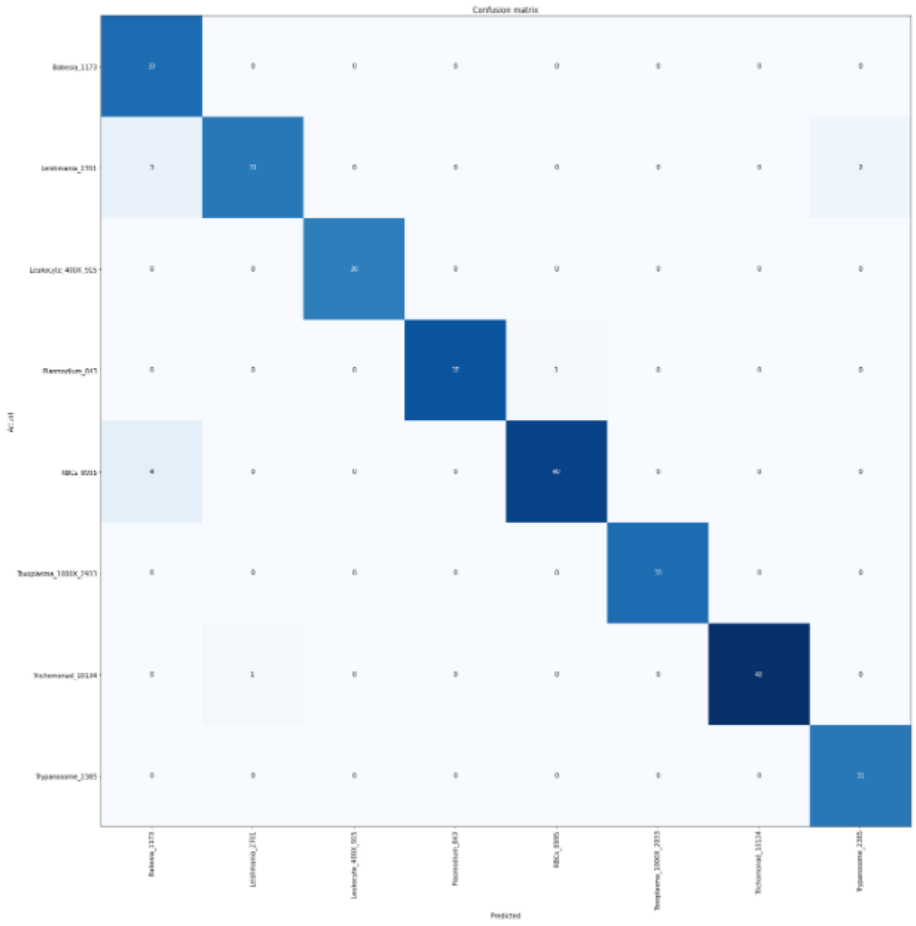

Figure 8: Confusion matrix of densenet 201

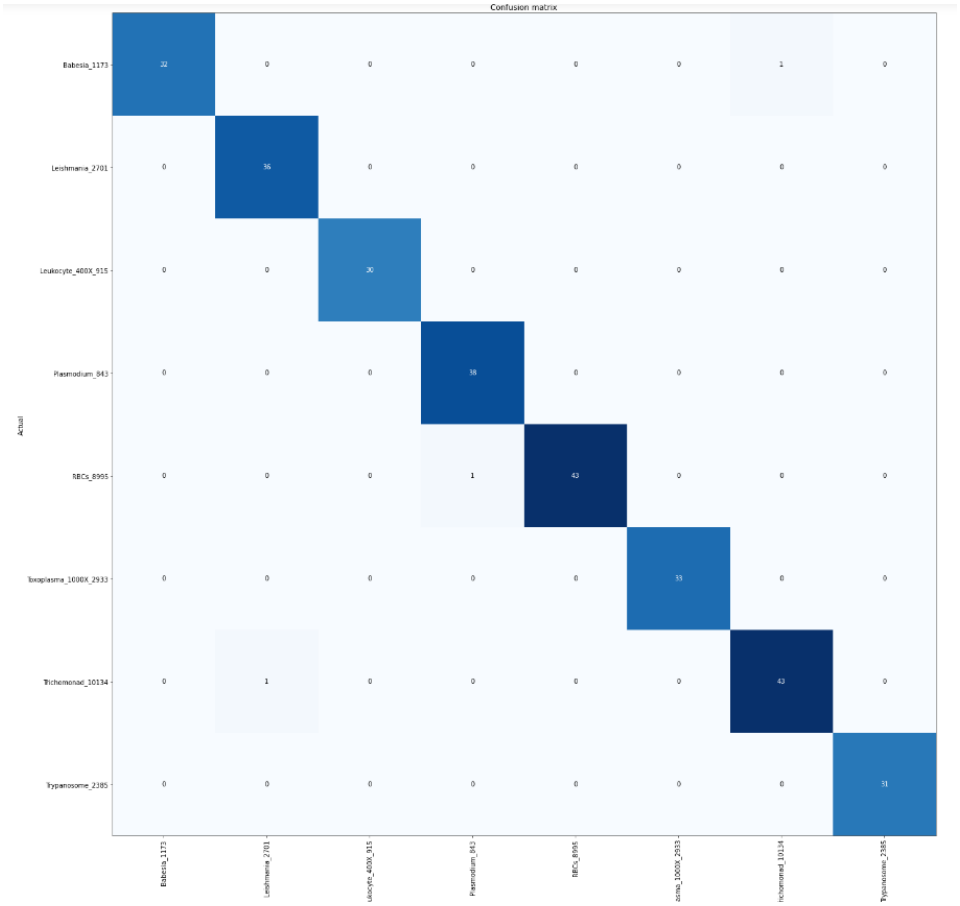

Figure 9: Confusion matrix of densenet 201 with the best learning rate

4.4 *VGG16*

Another model that is used is VGG16. In the training set, According to figure 20, VGG16 model has accuracy and f1 score of 0.968 which is good. Looking at the train loss it can be observed that the model had a poor performance in the train set. Additionally, the train loss is significantly bigger than valid loss which means the model is under fitting. Also the VGG16 model scored a really high roc auc score of 0.999 which means it classified positives and negatives really accurately. In the figure 21 it is seen that some parasites are confused by the model however, the amount of confusion isn't significant to change the learning rate. To conclude, the VGG16 has a high accuracy, f1 and roc auc score which means the method was good at classifying parasites. In the test set VGG16 predicted 281 out of 361 parasite images correctly with an accuracy score of 0.7229. VGG16 confused the Trypanosome parasite with the RBC parasite the most.

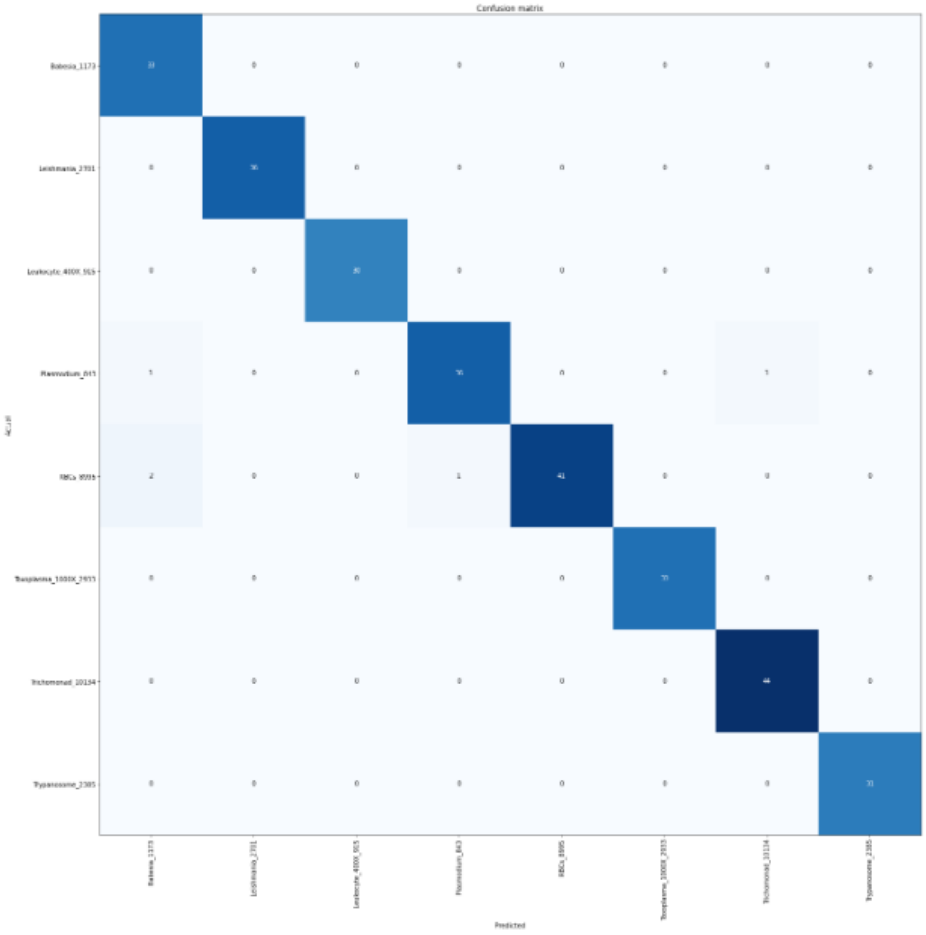

Figure 10: Results of VGG16

4.5 *Squeezenet1.1*

The last model that was used in the experiment is Squeezenet 1.1. In the training set Figure 22 shows that Squeezenet 1.1 had a high accuracy and f1 score of 0.986. Also, Squeezenet 1.1 have a high roc auc score of 0.999 which means positives and negatives were classified accurately. Also the data have under fitting since the train loss value is significantly bigger than valid loss value. To summarize, Squeezenet 1.1 performed really well with a high roc auc score, f1 score and accuracy score. However, there is significant under fitting problem.In the test set Squeezenet1.1 predicted 301 out of 361 parasite images correctly with an accuracy score of 0.877. Squeezenet1.1 confused the Babesia parasite with the RBC parasite the most.

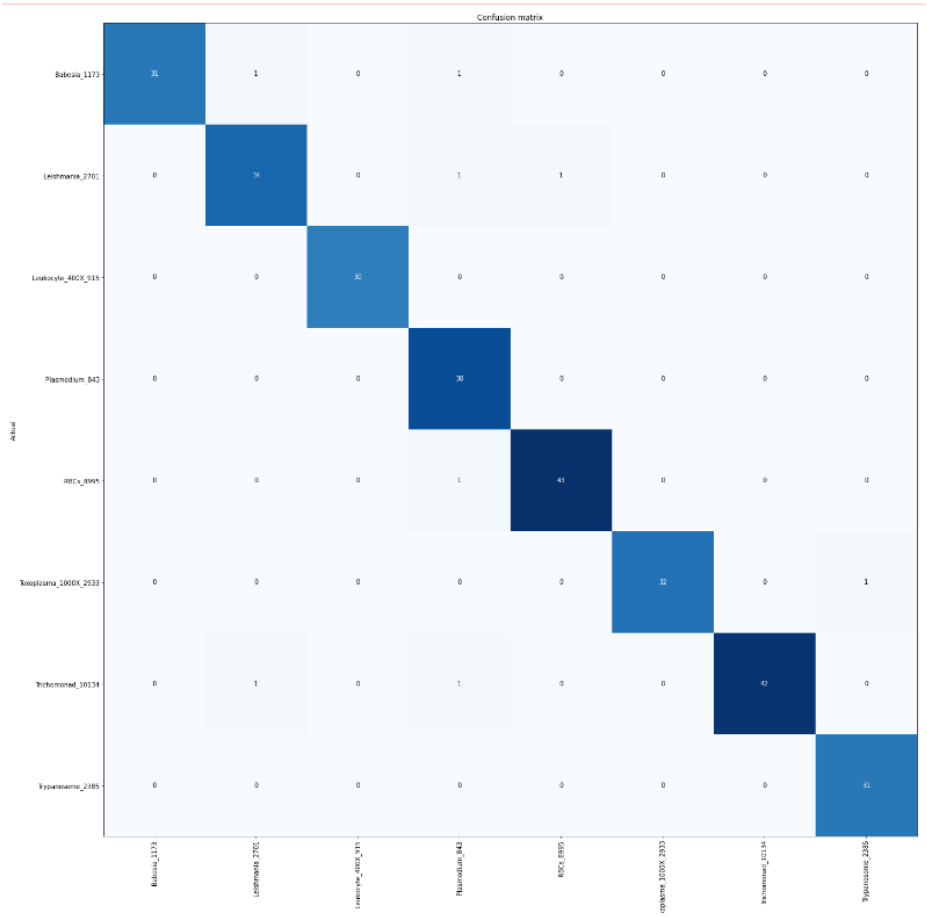

Figure 11: Confusion matrix of Squeezenet 1.1.

4.5.1 *Overview and Comparison*

In total there are eight pre-trained models and they are Alexnet, Resnet (34,50,101), Densenet(169,201), VGG16, and squeezenet1.1. In the training sets, all of the models are successful at detecting and classifying parasites and the models yield very good results. However, out of all the models, Densenet 201 performed the best achieving the highest accuracy and f1 score, and the Resnet 34 and VGG16 performed the worst. Interestingly Resnet 34 and VGG16 performed almost the same where the only differences are the models performances on the train and validation set. A common theme among the models is that all of them are suffering from underfitting. This shows that the models performed significantly better on the validation set and worse on the training set. Also, another reacquiring thing is that most of the models have a 0.999 Roc Auc score. This shows the model that we used was good at detecting positive and negative classes. As seen at Alexnet and Densene 201 Babesia and RBC parasites are the

most confused parasites. This makes sense since Babesia is kind of an RBC(Red Blood Cell) parasite. By looking at the Resnet and Densenet it can be concluded that if the model has more layers then it will perform better since more convolution layers mean more learning. However, the major setback of using models with many layers is that it increases computation time drastically. In the test set most of the models perform worse compared to the training set. An important founding was that in resnet 101 and VGG16 both miss labeled all of the trypanosome images as RBC images. Also, Trichomonad and Leishmania,RBC and Babesia were frequently miss labeled. The top 3 best performing models are densenet 169, densenet201 and Alexnet. And the worst performing model is both VGG16 and resnet101 with the same performance.

5 DISCUSSION

This research aims to develop a convolutional neural network that can identify and classify given parasite images. Therefore, multiple pre-trained convolutional neural network models were performed to find the best performing model. First of all, the answer to the research question, 'can we identify parasites using pre-trained convolutional neural networks?', is met with a yes and some of the models are good at it. In the training set, almost all of the models got a 0.999 Roc Auc score which means that the models were capable to classify positive class and negative class accurately and appropriately. The capability of convolutional neural networks to detect parasites was already mentioned in the literature. Additionally, to the current literature, the results also show that it can classify different parasites. When it comes to the sub-question, "which pretrained convolutional neural network should be used?", is a complicated answer. Since the choice depends on the preferences. In the case of best performance in the training set, densenet 201 is the best performing model with an accuracy score of 0.993. However, densenet 201 is a complex model therefore, it requires a lot of time and computational power. If a faster model was required Alexnet model would perform better since in the given case it only took the model 24 seconds to develop. Additionally, if the computational power is lacking the squeeze net should be preferred. To summarize, a model shouldn't be chosen solely on the performance rather it should be chosen according to the preferences or the requirements of the data set. However, in the training set, the best performing model was densenet 169 with 0.9778 accuracies. The densenet 169 was able to predict 329 images correctly out of 361 images. However, compared to the training set most of the models performed worse. For example, the resnet 101 model had an accuracy

value of 0.98 in the training set however, in the test set the model had the worst performance with an accuracy score of 0.78.

In the training/validation set all of the given models are successful at detecting and classifying parasites. While classifying all of the models performed well. A common thing in the models is that all of them showed a high ROC AUC score and all of the models had a fitting problem. This underfitting problem is caused by the training data performing poorly. The underfitting problem could be caused because of the size of the user data set. Since the data set has a small size therefore less amount of data might be causing this fitting. The densenet 201 had the best performance while VGG16 and resnet 34 shared the worst performance. Even though, VGG16 and resnet 34 had the worst performance still both of the models performed well and had significantly good scores. In the testing set, all of the models have misclassified the same parasites at least three times which was our boundary to adjust the hyper parameter. The models that had a hyper parameter adjustment(densenet 201, Alexnet) performed better than most of the models in the test set. In both of the sets, VGG16 performed the worst both of them. One interesting finding is resnet 34 performed worse on the training set compared to resnet 101. Yet, on the testing set the resnet 34 outpreformed the resnet 101.

In the training/validation set only major misclassification accrued on the Alexnet and Densenet 201 models. The Alexnet model confused Babesia as RBC. The Densenet 201 model also like Alexnet confused RBCS as Babesia. This confusion is understandable since Babesia is an RBC parasite. In the testing set, all models showed a significant level of confusion. An important finding was the VGG16 and resnet 101 model misclassified all of the trypanosome images as RBC. The trypanosome parasite has a more different and distinct look compared to RBCs which might be caused because of the size of the test set. Since in the training set both of the models were apple to distinct trypanosome and RBCs with no misclassification. Due to this trypanosome and RBCs are the most confused parasites with a total of 115 misclassification. Also, another finding is that the Trichomonad and Leishmania are confused frequently as well. There is a total of 31 misclassification cases with these parasites and the majority of these confusions occurred on VGG16 and resnet 101.

The research also gave interesting findings about the pre-trained models. For example, the squeeze net which is a model which is designed to be used by phones can perform as well as the other models which require more computational power. Additionally, the feature detection gives insight more info about which parasites could be confused the most. For example, in the training set both Alexnet and Densenet 201 mixed Babesia parasite

and RBC (Red blood cell) parasite. This confusion makes sense since Babesia is a type of RBC parasite.

6 CONCLUSION

The main question, 'can we identify parasites using pre-trained convolutional neural networks?', was the main idea behind this research. After employing pre-trained models to the dataset it turns out that convolutional neural networks are both an effective and efficient way to detect and classify parasites. Additionally, convolutional neural networks yield a lot of potentials and there is a lot of space for improvements. The study itself benefits from 8 different pre-trained CNN models. This way both an optimal model could be determined and a better understanding could be achieved by comparing the results of the models. In future research, the main idea should be based on building a new framework that could autonomously accurately detect and classify different parasites to achieve more specifications. The main rules should be followed while doing a classification task like this. First of all, the images should be represented accurately and a proper model should be picked for a better evaluation. Additionally, while implementing a model future approaches should focus more on building a proper model by feeding the model with more data. Therefore, the data would be able to learn more from the training and give a more realistic and accurate representation. Additionally, concerning the current findings and the literature while doing a follow-up study without digression from the topic working with a data set that has more unique parasites is suggested. By doing this, the researcher could classify more parasites and undertake more inclusive classification tasks. Also, this way the data could reveal interesting relations and outcomes. Another, follow-up study could be using both machine learning and deep learning together to achieve higher performance like the one achieved in the research of Fuhad et al(2020). Additionally, using the idea from Saeed, Jabbar (2017), and Fuhad et al(2020) pre-trained squeeze net model could be used on a smartphone so that the phone both can identify and classify different parasites. the result of this study could be used to improve ideas such as using a bigger part of the dataset, trying different pre-trained algorithms, or training a CNN from scratch to improve the performance. With the broadness of CNN, a lot of new innovative ideas or questions could spark. To conclude, using convolutional neural network models for the diagnosis of the disease have a promising future and a lot of potentials. The ability to detect and diagnose certain diseases could save a lot of lives and resources.

REFERENCES

Momčilović, S., Cantacessi, C., Arsić-Arsenijević, V., Otranto, D., Tasić-Otašević, S. (2019). Rapid diagnosis of parasitic diseases: current scenario and future needs. Clinical Microbiology and Infection, 25(3), 290-309.

Serban, G. (2019). Future prospects in the treatment of parasitic diseases: 2-amino-1, 3, 4-thiadiazoles in leishmaniasis. Molecules, 24(8), 1557.

Saeed, M. A., Jabbar, A. (2018). "Smart diagnosis" of parasitic diseases by use of smartphones. Journal of clinical microbiology, 56(1), e01469-17.

Talapko, J., Škrlec, I., Alebić, T., Jukić, M., Včev, A. (2019). Malaria: the past and the present. Microorganisms, 7(6), 179.

Yadav, S. S., Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. Journal of Big Data, 6(1), 1-18.

Li, Sen; Zhang, Yang (2020), "Microscopic Images of Parasites Species ", Mendeley Data, V3, doi: 10.17632/38jtn4nzs6.3Momčilović, S., Cantacessi, C., Arsić-Arsenijević, V., Otranto, D., Tasić-Otašević, S. (2019). Rapid diagnosis of parasitic diseases: current scenario and future needs. Clinical Microbiology and Infection, 25(3), 290-309.

Serban, G. (2019). Future prospects in the treatment of parasitic diseases: 2-amino-1, 3, 4-thiadiazoles in leishmaniasis. Molecules, 24(8), 1557.

Saeed, M. A., Jabbar, A. (2018). "Smart diagnosis" of parasitic diseases by use of smartphones. Journal of clinical microbiology, 56(1), e01469-17.

Talapko, J., Škrlec, I., Alebić, T., Jukić, M., Včev, A. (2019). Malaria: the past and the present. Microorganisms, 7(6), 179.

Yadav, S. S., Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. Journal of Big Data, 6(1), 1-18.

Li, Sen; Zhang, Yang (2020), "Microscopic Images of Parasites Species ", Mendeley Data, V3, doi: 10.17632/38jtn4nzs6.3

Das, D. K., Ghosh, M., Pal, M., Maiti, A. K., Chakraborty, C. (2013). Machine learning approach for automated screening of malaria parasite using light microscopic images. Micron, 45, 97-106.

Kornblith, S., Shlens, J., Le, Q. V. (2019). Do better imagenet models transfer better?. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2661-2671).

Pardamean, B., Cenggoro, T. W., Rahutomo, R., Budiarto, A., Karuppiah, E. K. (2018). Transfer learning from chest X-ray pre-trained convolutional neural network for learning mammogram data. Procedia Computer Science, 135, 400-407.

Umer, M., Sadiq, S., Ahmad, M., Ullah, S., Choi, G. S., Mehmood, A. (2020). A novel stacked CNN for malarial parasite detection in thin blood smear images. IEEE Access, 8, 93782-93792.

Benbihi, A., Geist, M., Pradalier, C. (2019). Elf: Embedded localisation of features in pre-trained cnn. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 7940-7949).

Koonce, B. (2021). SqueezeNet. In Convolutional Neural Networks with Swift for Tensorflow (pp. 73-85). Apress, Berkeley, CA.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362.

Howard, J., Gugger, S. (2020). Fastai: a layered API for deep learning. Information, 11(2), 108.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164.

Ricciardi, A., Ndao, M. (2015). Diagnosis of parasitic infections: what's going on?. Journal of biomolecular screening, 20(1), 6-21.

CDC - parasites - about parasites. (2022, March 21). Retrieved May 17, 2022, from https://www.cdc.gov/parasites/about.html Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science engineering, 9(03), 90-95.

Pomari, E., Piubelli, C., Perandin, F., Bisoffi, Z. (2019). Digital PCR: A new technology for diagnosis of parasitic infections. Clinical Microbiology and Infection, 25(12), 1510-1516.

Fuhad, K. M., Tuba, J. F., Sarker, M., Ali, R., Momen, S., Mohammed, N., Rahman, T. (2020). Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application. Diagnostics, 10(5), 329.

CDC - Babesiosis. (2020, March 31). Retrieved May 20, 2022, from https://www.cdc.gov/parasites/babesiosis/index.html

CDC - Babesiosis - diagnosis. (2018, July 17). Retrieved May 20, 2022, from https://www.cdc.gov/parasites/babesiosis/diagnosis.html

CDC - Babesiosis - disease. (2018, May 17). Retrieved May 20, 2022, from https://www.cdc.gov/parasites/babesiosis/disease.html

CDC - Malaria - about malaria - faqs. (2022, March 22). Retrieved May 20, 2022, from https://www.cdc.gov/malaria/about/faqs.html

CDC - toxoplasmosis. (2018, August 29). Retrieved May 20, 2022, from https://www.cdc.gov/parasites/toxoplasmosis/index.html#:~:text=Toxoplasmosis%

20is%20considered%20to%20be,the%20parasite%20from%20causing%20illness.

CDC - toxoplasmosis - general information - frequently asked questions (faqs). (2020, September 03). Retrieved May 20, 2022, from https://www.cdc.gov/parasites/toxoplasmosis/gen_info/faqs.html

CDC-Std Facts - Trichomoniasis. (2022, April 25). Retrieved May 20, 2022, from https://www.cdc.gov/std/trichomonas/stdfact-trichomoniasis.htm

CDC - african trypanosomiasis - general information - east african trypanosomiasis faqs. (2022, February 25). Retrieved May 20, 2022, from https://www.cdc.gov/parasites/sleepingsickness/gen_info/faqs.html

APPENDIX A

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	1.649753	0.292859	0.083045	0.916955	0.915876	0.994912	00:18

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.299197	0.145156	0.034602	0.965398	0.965530	0.997472	00:24

Figure 12: Results of Alexnet

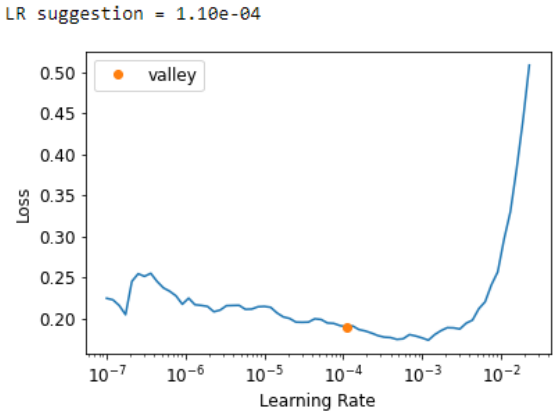

Figure 13: learning rate plot for Alexnet

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.192959	0.104239	0.017301	0.982699	0.982669	0.998214	00:14

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.147089	0.102317	0.017301	0.982699	0.982700	0.998008	00:24
1	0.161717	0.095289	0.013841	0.986159	0.986152	0.998301	00:24
2	0.174689	0.094877	0.013841	0.986159	0.986152	0.998147	00:24

Figure 14: Results of Alexnet with the best learning rate

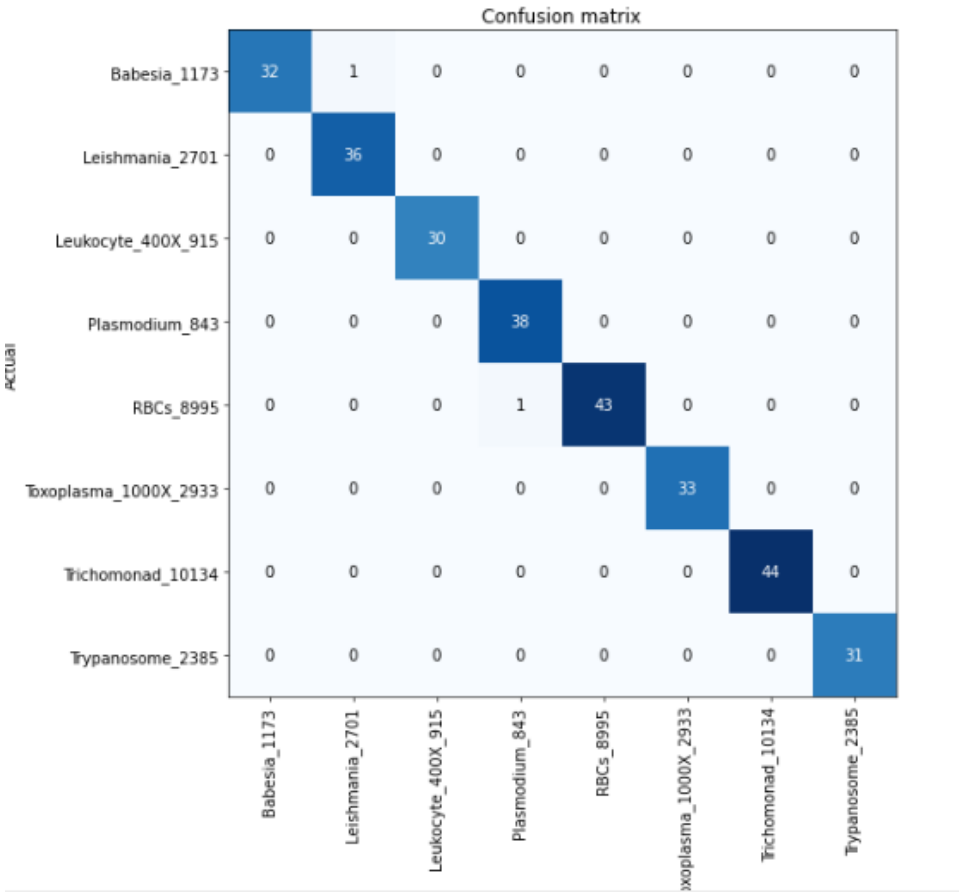

Figure 15: Confusion matrix of Alexnet with the best learning rate

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	1.795875	0.396080	0.138408	0.861592	0.859086	0.994392	01:54

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.226270	0.088972	0.031142	0.968858	0.968907	0.999362	02:21

Figure 16: Results of ResNet 34

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	1.070760	0.381541	0.100346	0.899654	0.906993	0.994075	03:59

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.130848	0.042680	0.020761	0.979239	0.979322	0.999854	04:47

Figure 17: Results of ResNet 50

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	1.008101	0.169058	0.055363	0.944637	0.945665	0.998449	06:40

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.133625	0.051160	0.017301	0.982699	0.982712	0.999710	08:07

Figure 18: Results of ResNet 101

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	1.155573	0.420630	0.138408	0.861592	0.862193	0.996574	08:11

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.122280	0.062816	0.024221	0.975779	0.975705	0.999891	07:14

Figure 19: Results of densenet 169

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	1.121379	0.343587	0.131488	0.868512	0.867849	0.997070	11:23

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.122233	0.113247	0.038062	0.961938	0.962248	0.999346	20:13

Figure 20: Results of densenet 201

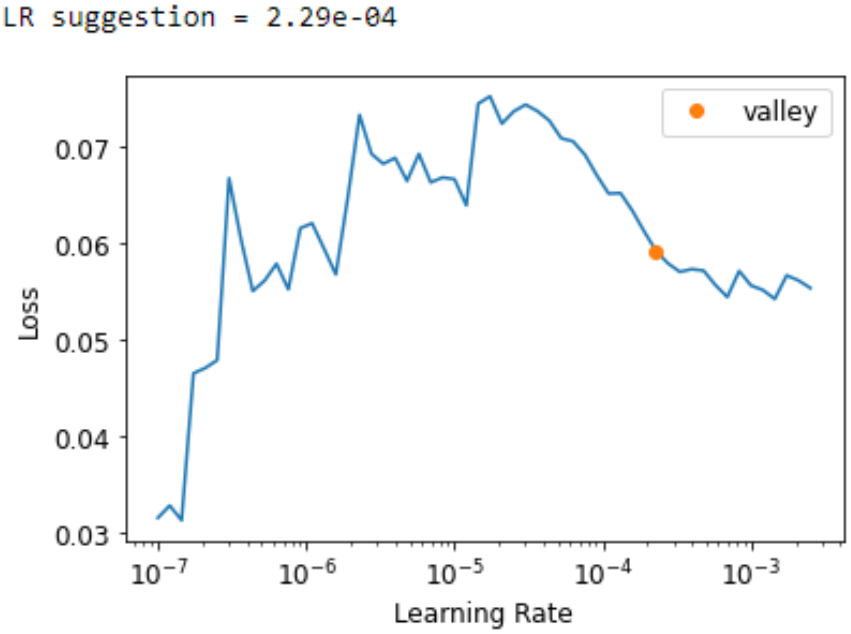

Figure 21: learning rate plot for densenet 201

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.067818	0.057504	0.020761	0.979239	0.979092	0.999686	18:51

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.049044	0.034896	0.006920	0.993080	0.993080	0.999834	35:42
1	0.038435	0.023126	0.006920	0.993080	0.993080	0.999897	38:46
2	0.045456	0.017523	0.006920	0.993080	0.993080	0.999936	49:44

Figure 22: Results of densenet 201 with the best learning rate

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	1.685958	0.416662	0.128028	0.871972	0.870850	0.995249	05:49

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.277816	0.085721	0.031142	0.968858	0.968938	0.999732	08:23

Figure 23: Results of VGG16

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	1.453932	0.097133	0.024221	0.975779	0.975945	0.998556	00:22

epoch	train_loss	valid_loss	error_rate	accuracy	f1_score	roc_auc_score	time
0	0.188468	0.037109	0.013841	0.986159	0.986200	0.999871	00:49

Figure 24: Results of Squeezenet 1.1.