Assessing Prediction Outcomes of Convolutional Neural Networks Trained on Parasite and Healthy Cell Images for Sickle-Cell & Malaria

: Abstract

To address the implications of using a CNN, a machine learning model was trained with parasite (positive) and healthy (negative) cell images of each respective disease, sickle-cell and malaria. In a world where the field of AI-assisted diagnostics has experienced exponential growth in the past decade, the need for significant research is necessary. Through this project, the goal is to have two working CNN models (one model per one disease) that can accurately predict if a human has sickle-cell or malaria based on positive and negative cell image data sets. The initial hypothesis is as follows, if a CNN model is trained on positive and negative images for both diseases, it can accurately predict the presence of both with 90% accuracy. Both datasets came from Kaggle, but the Malaria data set originated from NIH datasets and the sickle-cell data set images were produced from a research study located in Uganda, Africa. Preprocessing steps included resizing images to a standard 255 by 255 pixels and color standardization to ensure the model is not skewed based on unintended biases. The model was mainly trained through a random forest algorithm, a supervised learning mechanism, and was split into a 10% testing set, 80% training set, and 10% validation set. These proportions differed when running ablation tests, but this was the standard testing split done. Evaluation techniques include detailed confusion matrixes, accuracy curves, and loss curves to interpret this model's results extensively.

The sickle-cell model had a higher percentage accuracy in all three types of split tests done, the 80/10/10, 80/20 and 90/10 when compared to the malaria model. It also had a smaller percent error, for both false negatives and false positives with a better fit accuracy and loss curve (malaria was much more overfit and had cases of chaotic loss during the first 0 to 2 epochs). From the results of the model, it gives strong evidence towards the claim that the CNN model was more productive for the sickle-cell model, as compared to the malaria model. More research needs to be done to confirm this fact, but it is a trend that can be seen throughout the study. Researchers and clinicians could use these models to not only detect sickle-cell/malaria in

patients with a simple image but can see if there are specific characteristics that pertain to one disease in particular. Although this research did not use a heat map, it is a future study that can be done to see if there are patterns in the physical characteristics of the cells respective to each disease. The above information can be synthesized to understand machine learning's larger role within disease diagnosis and its feasibility within a certain scope.

Section 2: Introduction

At first glance, malaria and sickle-cell are quite disparate disease types. Sickle-cell is a genetic disease caused by mutations in the B globin gene, which results in "sickle" (deformed) hemoglobin production. Its physical effects include acute & chronic pain to the body, intense organ damage, and a median life expectancy of 43 years in the United States (Cafasso, 2019). On the contrary, malaria originates from Anopheles mosquitoes, parasites that transmit the disease from one person to another. Hence, countries prone to warmer climates that house these mosquitoes, such as India or China, are more likely to demonstrate malaria cases (Brandow and Liem, 2015). Sickle-cell develops rapidly in the human body, making it a high morbidity and early mortality type of infection. In sub-Saharan Africa and other low-income countries, there has been a growing burden of sickle cell disease. Due to the rapid growth in overall birth rates, the number of infants possessing the B Globin gene mutation increased by 453,000 (13.6%) in 2020 to 515,000 (13.7%) in 2021 (Osei and McGann, 2023).

Due to this increase in infections, diagnosis is a crucial step in resolving many disease application issues. In recent decades, disease diagnosis has become increasingly expensive, time-consuming, and inaccurate. To counteract this, image-based modeling and rendering methods have come into fruition. They enhance diagnostic accuracy, streamline workflows, reduce interpretation times, and improve patient outcomes. Potential use cases include visualizing internal structures and detecting physical abnormalities. Deep learning, specifically Convolutional Neural Networks (CNNs), have upended the world of medical imaging due to its ability to feature extract with ease and uncover subtle patterns unnoticeable to the naked eye. Both sickle-cell and Malaria are often used as disease data sets in CNN models due to their similarity of altering the red blood cell shape. Although one is a genetic disease and the other is

acquired from other species, there is value in studying them together through image-based research to consider potential treatments based on their shared characteristics and overall outcomes (Brandow and Liem, 2015).

The most common diagnostic tool for Malaria is blood smear microscopy (sample of blood examined under a microscope, used for nearly eighty years. This process relies heavily on the reliability of the technicians to take accurate blood smears, meaning the diagnostic results' variability could increase. Without a skilled, reliable technician the results render obsolete. Stained smears are also not uncommon, causing false positive and negative diagnoses results and even longer disease resolution timelines (Grimberg, 2015). Despite the lack of accuracy and reliability, many developing countries still value blood smear microscopy as it is extremely cost effective. Data from rechecking malaria blood slides at general hospitals in 2018 reported 238 out of 844 imported cases (28.2%) that were not correctly diagnosed thus stressing the urgent need for a more accurate diagnosis tool (Li et al., 2020). Aside from more apparent cases (avascular necrosis on imaging or leg ulcers on an exam), there are no reliable diagnostic tests to confirm the presence of acute or chronic pain specific to Sickle-Cell, which makes disease diagnosis especially challenging. There are several ways to go about the diagnosis of a genetic disease: a physical examination, analyzing personal medical history, and conducting laboratory tests (including genetic testing). Even though both sickle-cell and Malaria are seen as disparate diseases, attempting to find the connection in their diagnosis could be incredibly beneficial. Their shared challenges such as variability in staining/quality of blood smears and under-resourced lab infrastructure make scalability extremely difficult. By studying diagnosis characteristics such as cell shape, size, or location it allows for a wider interpretation of results and gives potential information on their shared treatment possibilities (Arishi et al., 2021).

Machine learning as a diagnosis tool has been applied to disease applications for several years. It is a method researchers use to solve public health issues that already have previously acquired data as without prior data it is almost impossible to make a machine learning model successful. Automated blood cell detection and classification is one example that is topical to this current study. Identifying blood cell types is crucial to diagnosing hematological diseases as these infections originate from a blood source. Even though sickle-cell anemia and malaria,

which although are not directly hematological, are still heavily impacted by red blood cells. The first step of intervention is the microscopic examination of blood cells to identify any abnormalities, or unusual characteristics present at the scene. Recently, the physical act of manual detection has come to a stop, and many researchers have opted in for automated medical imaging processes such as deep learning/machine learning. It transformed the field of blood cell detection, enhancing the precision and speed of diagnostics of any other physical process (Sazak, Kotan, 2024). Another successful example of medical imaging includes cancer detection (Kononenko, 2001). Mammograms are a tool used to screen for breast cancer, to ensure those with increased risk for developing a particular type of cancer are able to be detected as early as possible. Imaging can also help determine how advanced the cancer is and what its current stage is, a step that is almost impossible to do without deep learning technology ("Cancer Imaging Basics"). Many life-threatening conditions, such as cancer or heart disease, develop slowly over time. These diseases are usually easier to treat or manage if caught in their initial phases, including sickle-cell and malaria. Machine learning can learn distinguishing features of red blood cell shapes and infected parasite cells in sickle-cell and malaria, which can make for more accurate and timely diagnosis that reduces dependence on highly trained, specialized personnel.

To solve problems presented throughout, it is important to consider this essential question: Can convolutional neural networks trained on parasite and healthy red blood cell images accurately distinguish between sickle cell disease, malaria infection, and healthy cells, and what preprocessing steps best improve classification performance? Evaluating this question will not only offer clinicians decision support and reduce delays due to contaminated samples, but it tackles two major blood-related diseases simultaneously. Additionally, this study will employ multiple machine learning algorithms including Random Forest models and CNNs, to determine what type fits best depending on their use case (Lei et al., 2020). By training two disease-specific models with a shared pipeline, we can effectively conduct malaria parasitemia detection and sickle-cell morphology classification using the same framework. This dual-disease design demonstrates the pipeline's reusability across related smear-based diagnostics, enabling a like-for-like comparison of pre-processing and modeling choices across diseases which proves necessary in settings where both malaria and SCD screening co-exist.

Section 3: Prior Research

Limitations of blood smear microscopy contributed to the failure of many crucial government programs, including the 1950s and 1960s WHO Global Programme to Eliminate Malaria. As this process relies heavily on the reliability of the technician, it causes unwanted variability in the results. Several diagnostic technologies have come about to tackle these specific issues. Flow cytometric devices contribute to malaria diagnosis by relying on the visualization of parasites on stained blood slides. False positive rates in microscopy of up to 36% and false negatives as high as 18% are still unfortunately common, even with the use of artificial intelligence analysis. The use of radioactive hypoxanthine was also developed, made to reduce the subjective nature of microscopic assays. A challenge that comes with using hypoxanthine is its need for radioactivity and its difficulty in differentiating when certain substances are made by human cells or mosquito parasites (Grimberg, 2011). Specifically for sickle-cell, current techniques to diagnose and monitor SCD include similar drawbacks. A completed blood cell count is done to characterize the different types of anemia and figure out their hematological parameters (measures used to evaluate blood characteristics). Next, a peripheral blood smear (done after spotting abnormality in the automation counts, landmark of hematological evaluation) is completed. Yes, it is simple, rapid, and relatively inexpensive but the small availability of trained pathologists is another factor.

Every positive tends to come with a negative consequence that is difficult to counteract. For this reason, digital image processing has made a big difference in modern diagnostics processes. Three examples include: Region-based segmentation, where an image is split into different parts based on color, texture, or other visual features; Edge-based segmentation, to figure out where one object ends and another begins; Morphological clustering, which grouped pixels based on their shape and their location on the plane. To successfully use deep learning techniques described above, NIH malaria cell images and publicly available sickle cell blood smear datasets became potential candidates for comprehensive data. To this day, if a researcher does not have the ability to collect their own data, they rely on free digital datasets that have already been tested for accuracy (i.e NIH or Kaggle).

Malaria detection through machine learning has become a recently common method to obtain higher accuracy in diagnosis. Deep learning model architecture is designed for the task of detecting malaria from images of blood samples. It uses an encoder/decoder scheme for image segmentation tasks and involves two separate neural networks to process the input image and generate an output segmentation map.

The inclusion of residual layers can improve the model's performance by allowing it to grasp complex relationships quicker within the data and make more accurate predictions. Residual layers are a type of layer that forces the model to learn residual mapping (difference between observed and predicted values) rather than attempting to learn mapping from scratch. The segmentation model described above achieved a 97.1% per-pixel accuracy, when using a blood smear data set collected from the arm using a syringe with standard procedure. Not only was the accuracy relatively high, but the evaluation time per image was around 22ms meaning the model was lightweight enough to rapidly run on even less powerful devices. A second algorithm used within this same study was the "cell counting algorithm." In short, by comparing the output of the neural network within the original inputs, it is possible to see how accurately the network is able to classify various types of cells in the samples. After 1000 epochs of training, the network achieved a 97.1% per-pixel accuracy, the same as algorithm one. Overall, this method was incredibly efficient and has low-resource deployment, making it accessible to clinicians around the world, not only in first world countries (Thakur et al., 2015).

A common theme in many malaria parasites images is the use of thin blood smears rather than thick blood smears. WHO has recommended that microscopic diagnosis be only performed on thick smears as it gives a higher concentration of the sample. A low-cost alternative was presented for rapid malaria screening using a modified YOLOv3 detection algorithm. The YOLO algorithm utilizes a single neutral network to a full image split into multiple grid sections. Then, objects within their respective grid cells are detected by the grid cells themselves. Performance of this algorithm was tested on two data sets, collected by a microscope camera and mobile phone camera. The first model achieved 99.07% accuracy and the second 97.46% accuracy. Although the mean average precision of the model was similar to the rate of a human technician, using the modified model was much more efficient (Chibuta and Acar, 2025).

Machine learning processes for sickle-cell were widely similar, in that many studies covered similar algorithms but produced slightly disparate results. Data augmentation was used in a 2021 study conducted by Dr. Tegenshe R to increase the size of the training data using the existing data available. This also reduces the chance of over-fitting the model by ensuring that the images created are not duplicate of the existing ones already in the data set. The next pre-processing step was histogram equalization to adjust the contrast in the image by changing the histogram's intensity distribution. In order to evaluate the model on parasite localization (estimating the parasite's location accurately), it is necessary to determine how well our model predicts the location of all types of parasites. To achieve this, a bounding box was drawn around the parasite and the overlap ratio between the predicted bounding box and the ground truth bounding box (actual location) was evaluated. When looking at the images with the lowest precision from both data sets, in both cases, the model was initially confident on the parasite's location within the bounding boxes. This can be attributed to missed labelling of the bounding boxes done by the researchers. Even minor errors such as this have a massive impact on the overall performance of the model in the test. As a result, the precision was 0.056, the accuracy was 0.055 and had an accuracy value of 99.07%. Thus, the model should be able to accurately classify with confidence as to whether a patient has malaria infection or if they are negative. As sickle-cell is a much less common disease to Malaria in countries outside of the United States, it is challenging to obtain quality data sets. Conversely, malaria is rare in the United States making it a bottom priority for many domestic lab corporations and, as a result, little progress has been made on increasing its diagnosis efficiency (Hoyos, 2024).

Although many studies focus on the treatment and diagnosis of Malaria and Sickle-cell separately, a combined pipeline method will be used further. Most importantly, by including Grad-Cam to evaluate our results and demonstrate the exact location at which the model is designating a cell image as "Positive" or "Negative" will give us potential insights helpful in understanding both diseases and their diagnosis. Many of the preprocessing steps and the backbone of CNN models do not differ for both diseases as, if the goal of the research does not change, neither would the type of evaluation done on them. That said, the specific physical characteristics of each type of data for sickle-cell and Malaria do differ. Sickle-cell diseased cell

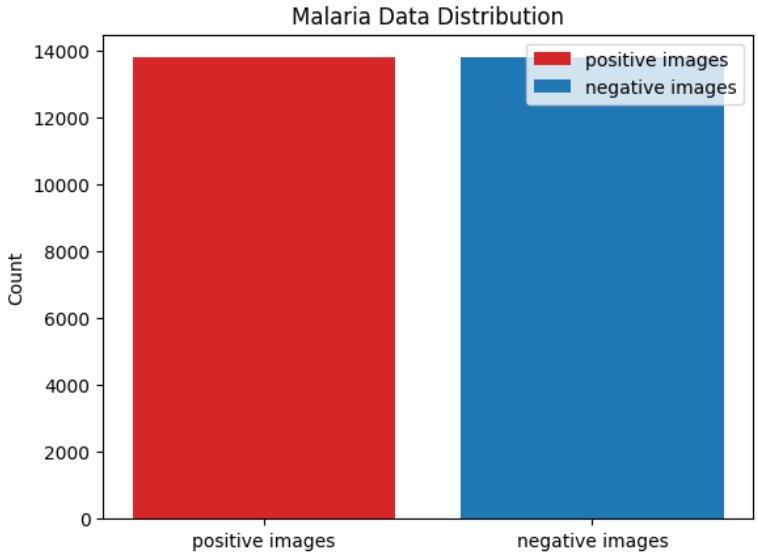
images clearly demonstrate the sickling of blood cells under low oxygen conditions, creating a Plasmodium falciparum infection. This process leads to damaged red blood cell membranes, making them leak essential nutrients that parasites need for survival. This infection may also cause a more severe version of malaria by binding to and blocking small blood vessel walls, leading to intense organ damage. In simple cell data sets, the stage at which the infection is at can rarely be detected without extremely advanced technology present. This quality is shared with the life cycle of malaria, which is similarly complex in nature. There are four main stages that a malaria parasite passes through including the initial infection by the mosquito bite, liver stage (infects liver cells), blood stage (ruptures blood cells), and transmission to the mosquito (gametocytes ingested by the mosquito). As one can see, the developmental stages cannot be predicted by an image-based classification model such as a Convolutional Neutral Network, but it can provide us information on what specific characteristics of the image categorize them Positive or Negative for the disease (Hoyos, 2024).

Section 4: Methodology

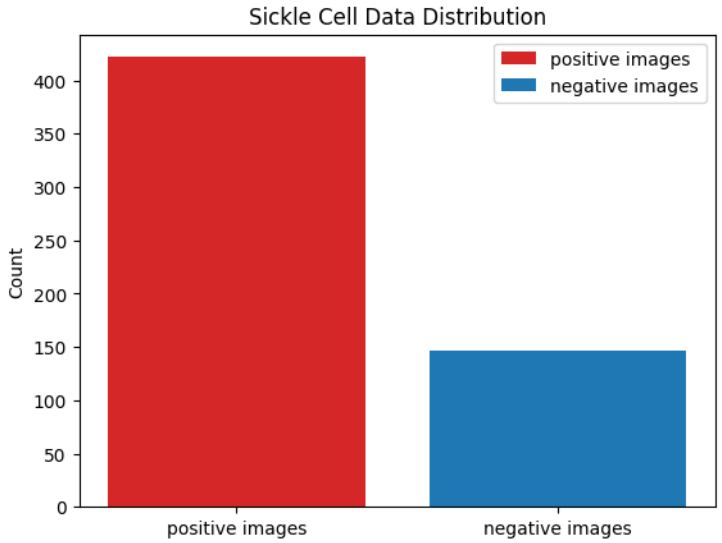
The overall research design can be described as two exploratory CNN models designed to test out various quantitative measurements to determine how they affect accuracy, precision, and f1 scores. Essentially, a quantitative, experimental study using two separate CNN classifiers, one for malaria and one for sickle-cell disease, was used to evaluate how pre-processing choices and network architecture affect classification accuracy, precision, recall and F1. Using a CNN was the right choice as in a real-world situation, clinicians would only be able to confirm whether a patient is infected or not with the use of an analyzed image of blood smear. As two models were created, one for malaria and one for sickle-cell disease, it was important to find two data sets that included clear and accurate images. The sickle-cell disease data set were images collected from the Teso region in Uganda, with picked samples from the Kumi and Soroti district specifically (NIH Microscopic Data Image Set). 140 patients provided their blood samples which were processed using two methods: field stains and leichman stains. Their microscopic images were cleaned, processed, and placed into this data set which includes 422 positive (sickle cell) images and 147 negative images. Note, this data set included 122 unclear images that were not used in the training and testing of the model as they were either cropped, badly stained, or had an un-

usual color grading (Sickle-Cell Disease Blood Smear Dataset, Kaggle). The Malaria data set is simpler as they were obtained from the official NIH website, with a total of 27,558 images (both infected and uninfected).

The picture demonstrated below does not match the total number of images written above as even though a total of 27,000 images were provided in the Kaggle dataset, due to Google Collab's ram maximums and the time it would take for the GPU to keep up with the large number of images, it was more practical to choose a smaller set of images (for both positive and negative). Specifically, into a 80/10/10 split (80% training, 10% testing, and 10% validation). Numerically, 800 training samples, 100 validation samples and 100 test samples were used.



Similar to the malaria model, the sickle-cell data set was also stratified to ensure a balance between positive and negative images. The sickle-cell model used the 80/10/10 split, with 455 training samples, 57 validation samples and 57 test samples. Stratification is particularly useful for imbalance data sets, so it was important to use this as part of my data organization method. A random state of 42 was also instantiated to ensure reproducibility of the project. Originally, the picture below represents the number of total positive and negative images there were in the untouched dataset.



The pre-processing method was relatively limited as there was not an abundance of items that needed to be changed once the images were sorted. First, dimensions were resized to 255 by 255 pixels as it would make the picture large enough to see clearly without it taking up too much time and/or space when the model was running in real time. Next, color standardization occurred to ensure that all images had the exact same color balances, and the model would not pick up on attributes that actually had no defining impact on the results. While those were the two physical characteristics that were altered in the preprocessing, each image was also appended to an array that contained all usable images for each data set. This method was chosen as it was the only way to ensure that when deciding which images, the model would choose to test, they were as randomized as possible. Randomization occurred because the Google Collab interface, where both of these CNN models ran, did not have enough GPU to handle 27,000 malaria images at the same time. A potential risk of this is since there are fewer overall images in the training and testing section, it could have an impact on how accurate the model truly is. A general principle is that more data means a more accurate representation of one's model, meaning that this particular strategy may have a negative consequence to keep in mind.

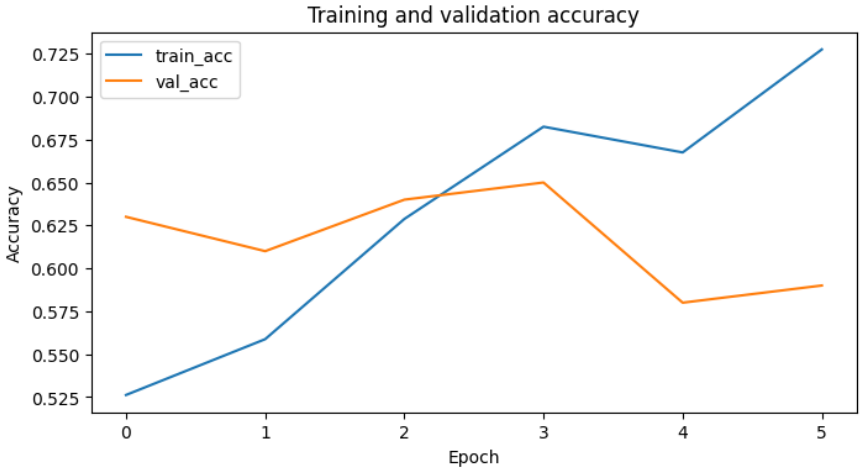
As machine-learning models are based on the accuracy evaluation of both models, the metric that would be most important to observe is accuracy. That being said, if the recall and F1 value are extremely low it would be something to consider when coming to an appropriate conclusion. All four metric types (accuracy, precision, F1, and recall) need to come together

harmoniously as accuracy may not be a good measure if the dataset is not balanced or if the negative and positive classes have a different number of images. A method that allows us to identify these evaluation metrics is a confusion matrix, a valuable tool to test out the performance of a disease classification model. It results in four key components: True Positive, True Negatives, False Positive, and False Negatives. For a sickle-cell-disease and Malaria disease classification model, a false negative would be the most harmful long term. If a patient were to actually be positive with either of those infections but be told by a medical professional they were in the clear, it could result in prolonged symptoms and in the worst cases could become fatal.

The two main libraries used were TensorFlow and Keras to build and train these machine learning models. TensorFlow supports large-scale applications with an option for GPU computing in Google Collab, making it an appealing option. Keras is a neural network that runs in addition to TensorFlow, making them a package most often. Both libraries support various model types in a much more streamlined manner, including the CNN model that is being discussed here. As all libraries were directly imported with a singular line of Python code, it is easily able to be reproduced in a different Google Collab space. What cannot be easily deciphered from online documentation of these libraries is the ethical nuances that come with them. As there is little information about the background of the chosen data sets in terms of their quality, the clinicians whose job it was to take the samples, or the timeline of when these microscopic images were taken, it is difficult to confirm their cleanliness and accuracy. Pointing out information gaps is important as it can point out potential errors or biases peeking out in the results. Similar to how stained or unclear microscopic images can affect the stability of data sets, the geographic region of where the images were taken can produce direct or indirect biases. For example, as the Malaria dataset was taken in Uganda, a third world country, the reported data is more often than not skewed to be Euro-centric or not considered due to the limited availability of skilled technicians. Regardless of if they are actually skilled or not, there are inherent biases still demonstrated (Silka, 2023).

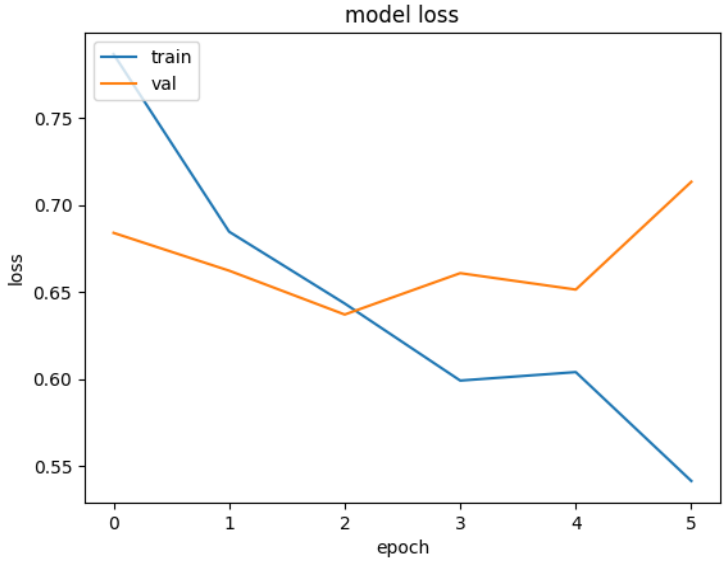
Section 5: Results

During training and validation, the Malaria model performed fittingly with the amount of preprocessing and external factors that affected its results. As the graph demonstrates, the accuracy is not increasing over time per epoch (for validation) due to potential overfitting in the validation curve (used to give an idea of how well the model is generalizing data). When the model becomes overly focused on the training data and does not perform well on unseen data, it can cause accuracy to decrease every epoch it has completed. Conversely, in the training curve (used to give an idea of how well the model is learning), accuracy is increasing at a different rate per epoch completed. From epoch 1 to epoch 2, accuracy changes by around 0.6 but from epoch 1 to epoch 2, accuracy increases by 0.5 indicating a non-linear rate of change.

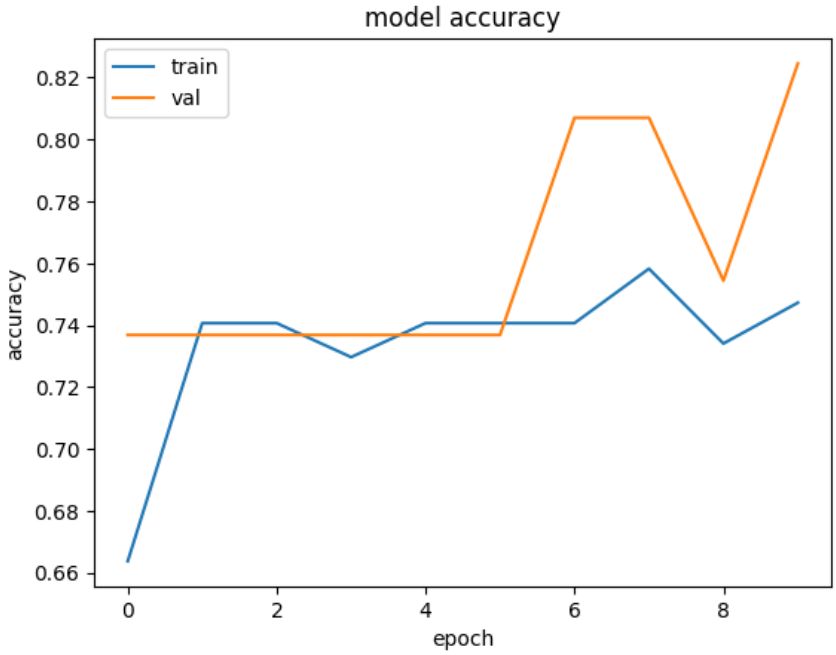


Other than accuracy curves, interpreting loss curves in the context of this model is also crucial to understanding the mechanics behind the program itself. There is chaotic loss (fluctuations per epoch) for both the training and validation curves, signaling that the training set was most likely not shuffled well enough, with too many oscillations between datasets occurring in the model. Although there is chaotic loss, there is a steady decrease proving the model is learning from data and is adjusting its parameters in real time to minimize errors. As loss is consistently decreasing for the training set, it adds to the idea that the model is well-performing. In addition, there is a significant gap between the two final loss values (hence, increasing the generalization gap that is bound to appear). Overfitting is again seen in the loss curve because

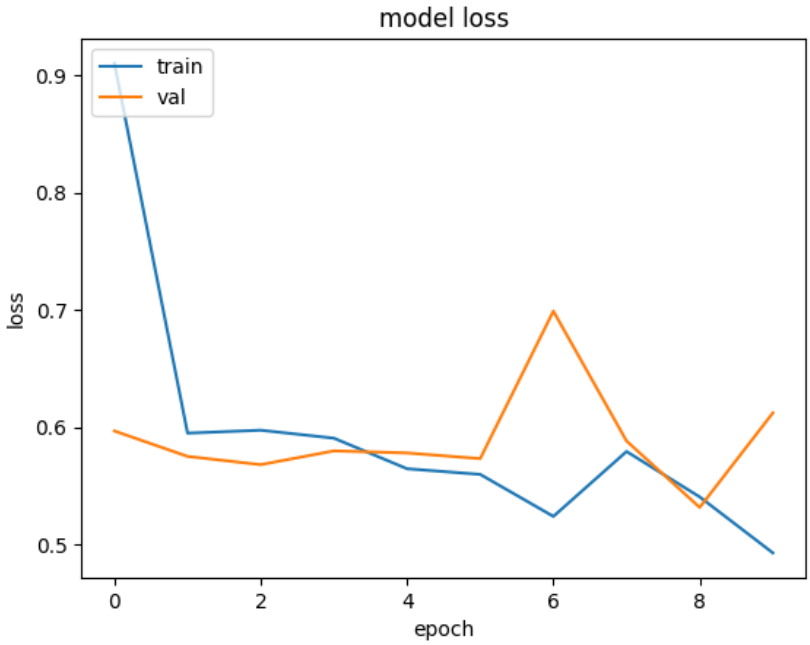
validation is higher at the end of 5 epochs than it was at the start of its run. There is a decrease from epoch 0 to epoch 2 of around 0.4, but loss does eventually get up to approximately 0.71.



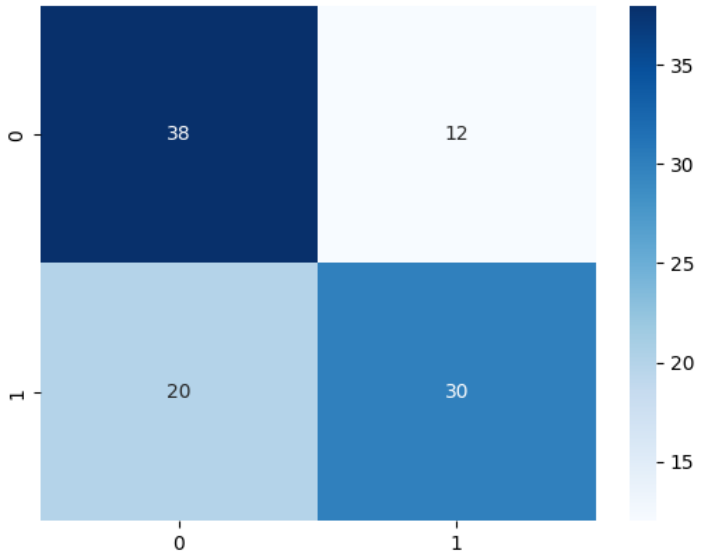
For the sickle-cell model, there were no differences in how the training and validation processes worked but there were changes to the results and their respective interpretations. The accuracy learning curves for both training and validations represent underfit as opposed to the overfit learning curves seen in the Malaria Model. It means the model was unable to fully learn the training data set due to the complexity of its overall.



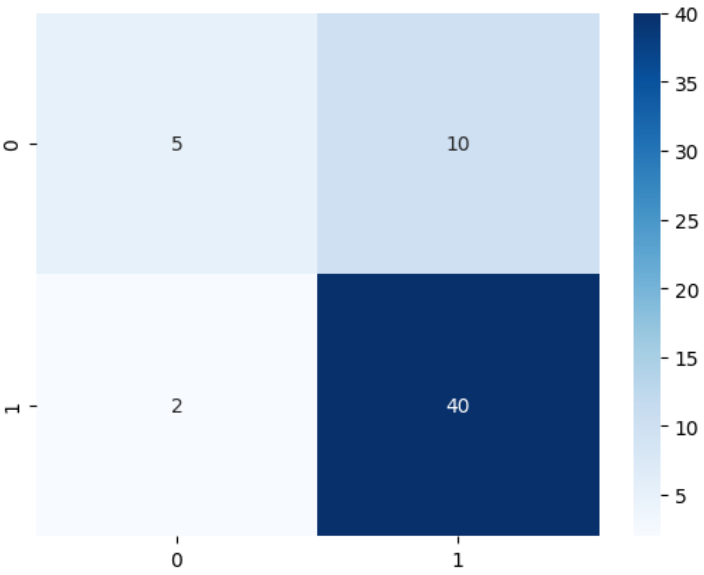
The loss curve for sickle-cell is similar to malaria in that there is chaotic loss seen as each epoch is completed. Although, the generalization gap between training and validation is higher, signifying that it is not the best representation of a good fit learning curve.



The final confusion matrix for the Malaria model can be found below, with true positives, false positives, false negatives, and true negatives.



The final confusion matrix for the sickle-cell-disease model can be found below with the same characteristics.



While the malaria confusion matrix had the most common error as false negative, in sickle-cell, false positives were the most common. Since pre-processing looks different depending on the type of model one is working with and the end goal in mind, it can act as a trial-and-error process to receive the best results. By choosing to try out different testing splits and how the images were resized, there can be a wider array of factors being tested. Results shown below are directly copy and pasted from the model to remain as integral to the study as possible.

Malaria	Sickle-cell
Original (80/10/10): Accuracy: 0.68 Precision: 0.6 Recall: 0.7142857142857143 F1 Score: 0.6521739130434783	Original (80/10/10): Accuracy: 0.7894736842105263 Precision: 0.9523809523809523 Recall: 0.8 F1 Score: 0.8695652173913043
With 20% testing and 80% training split: Accuracy: 0.7739130434782608 Precision: 0.9529411764705882 Recall: 0.7864077669902912 F1 Score: 0.8617021276595744	With 20% testing and 80% training split : Accuracy: 0.8070175438596491 Precision: 0.9764705882352941 Recall: 0.8058252427184466 F1 Score: 0.8829787234042553
With 10% testing and 90% training split: Accuracy: 0.7413793103448276 Precision: 0.9302325581395349 Recall: 0.7692307692307693 F1 Score: 0.8421052631578947	With 10% testing and 90% training split: Accuracy: 0.8421052631578947 Precision: 0.9761904761904762 Recall: 0.8367346938775511 F1 Score: 0.9010989010989011

When changing the testing split to 80/20 and 90/10 for training and testing, all four key components increased meaning there is a clear positive benefit to changing the splits for the

malaria model. Accuracy increased by 9 percent for 80/20 split, 6 percent for 90/10 split. For sickle-cell, when changing the testing split the accuracy, recall and f1-score also increased in value demonstrating the usefulness of trying out different training vs. testing splits.

Malaria	Sickle-cell
Original 255 by 255 pixels resizing (Malaria): Accuracy: 0.7739130434782608 Precision: 0.9529411764705882 Recall: 0.7864077669902912 F1 Score: 0.8617021276595744	Original 255 by 255 pixels resizing (sickle-cell): Accuracy: 0.8070175438596491 Precision: 0.9764705882352941 Recall: 0.8058252427184466 F1 Score: 0.8829787234042553
Resizing images to 300 by 300 pixels: Accuracy: 0.7652173913043478 Precision: 0.9294117647058824 Recall: 0.79 F1 Score: 0.8540540540540541	Resizing images to 300 by 300 pixels: Accuracy: 0.8421052631578947 Precision: 1.0 Recall: 0.8235294117647058 F1 Score: 0.9032258064516129

As seen on the sickle-cell side, when resizing each image from 255 by 255 pixels to 300 by 300 pixels, accuracy went up by approximately 4%, precision decreased by 3%, and both recall & f1 score changed by a negligible amount. Different splitting methods also changed metric values in a positive and negative way. For example, accuracy was the lowest in the original 80/10/10 split (68%) as compared to 77% and 74% for the 80/20 and 90/10 training and testing split (malaria model). Generally, a validation set is only used when there are hyperparameters set in place, such as standardization during pre-processing or a set number of layers in a neutral network. Since both models included these hyperparameters and more, it gave solid reasoning behind using the 80/10/10 split. But from the data, the evidence shows the

opposite: The splits with no validation set did better in all four metric types. The most frequent errors seen based on the confusion matrix are false positives over false negatives for sickle-cell. This is a better case scenario than false negatives as those can lead to prolonged symptoms with little treatment time and delayed care. Misclassifications of images (the action that leads to either false positives or false negatives) can stem from labeling errors, ambiguous images (images that contain multiple objects), and variability in appearance. In this situation, the most likely reason for misclassification comes down to variability in appearance. As preprocessing did not include standardization of lighting or angles, that would alter how the cell images appear, affecting how well the model can recognize the image. Conversely, the malaria model had higher amounts of false negatives. These are particularly unfortunate in the medical field because it promotes the spread of diseases if individuals are not taking precautions to protect others from their illness (when they are told they no longer possess the sickness). Adjusting the testing split to either a 80/20 or 90/10 could lead to a more accurate model, usable in low-resource settings, the overarching goal of the creation of this model. In addition, a better split could create lower rates of false negatives, the most harmful type of error.

Section 6: Conclusion & Discussion

Based on the presented hypothesis, the model will accurately predict the presence of sickle-cell disease or malaria with 80% accuracy. With the original 80/20/20 split, this prediction did not come true as the malaria model had an accuracy of 68% and the sickle-cell model had an accuracy of 78%. The sickle-cell model performed higher in terms of overall accuracy, which although is an important metric, precision is just as important in the medical field. If results are not replicable and/or not precise to the target, decreased productivity and resources are wasted each time a new study needs to be done to confirm previous results. A potential reason for a higher accuracy attributed to the sickle-cell model may be the objects seen in each diseased-cell image. In one sickle-cell image, there are dozens of tiny sickle cells dispersed throughout the lens while in one malaria cell image there is an abundantly clear malaria cell in the middle of the lens. This makes training the malaria model more difficult as the data does not have enough diversity in its appearance, creating overfit accuracy and loss curves for the validation set. The evaluation metrics described above do not represent a clean, one to one translation into a

clinically meaningful performance. A model, used to replicate a larger more complex process, is not guaranteed to work in a clinical setting with external factors such as technician abilities, items out of the researcher's control, coming in the way. The goal of a model is to almost reach the capability expectation of a real-life process but can't match the same level of uncertainty or variability of a real-world scenario. Moving forward with machine-learning models with caution is important as automated models should be used as decision support, not replacements for human clinicians.

In comparison with previous studies, the models include similar problems, while also being better fit in certain circumstances. Grimberg's 2011 study on blood smear microscopy with a machine-learning diagnostic tool showed a false positive rate of 36% with false negatives as high as 18% (Grimberg, 2011). The Malaria model had a false positive rate of 20% with a false negative rate of 12%. A decrease in both types of errors is seen, a good sign that this model can be used in hospitals where patients need to be treated efficiently. A possible reason for this difference in data could be aspects of methodology: Grimberg's study discusses the use of flow cytometric devices, which although have proved to be less successful in this newly created model, does not outright disprove its functionality. An example of a more unfortunate result is seen when reviewing Thakur's 2025 study, which created a segmentation model rather than a CNN visualization model. It achieved 97.1% per-pixel accuracy, almost 30% higher than the malaria model and 20% higher than the sickle-cell model. A potential reason for this stark difference could be the number of epochs all three models ran under: segmentation model ran 1000 epochs while malaria and sickle-cell both initially ran for 10 epochs and stopped once the loss became too high for the model to realistically move forward with. While the newly created model would still need refining and experimentation to fully understand its capability, it is still able to be somewhat more or less desirable, when compared against past studies that share the same research goal.

The impact of this model on global health can most be attributed to low-resource clinics as installing this model in second or third-world countries will provide easily accessible results in an efficient manner. When money is not in abundance, using digital models is especially useful as lab technicians do not need to have an immense amount of training. Because models are low

maintenance, researchers are more likely to conduct multiple early detection processes to patients to ensure faster treatment times and reduced overall costs. AI algorithms within this CNN model excel at recognizing patterns that indicate the beginning stages of illness, the phase at which disease detection is most important. Timely interventions and higher survival rates are two major pros of early detection mechanisms. Another scenario in which this model could be applied is in hospitals that use electronic health record (EHR) systems, which enable doctors to receive diagnostic feedback directly while evaluating patient history and current information.

An important aspect of model creations is to reflect on how one could go about improving it, whether it is through acquiring additional data, improving efficiency, etc. One method to strengthen this particular model is to conduct it on a platform that allows for more images to be used at the same time. Since Google Collab has limited GPU and it is not realistic to process 27,000 images at once, running the model on much more advanced software could allow one to use all available images at their disposal making their model have access to training data. Creating one synthesized model, rather than two separate models, where it predicts whether an image is parasitized or healthy for both diseases is another way to test out the variability in the model. The program behind the model needs to change but it may make the model more applicable in a variety of situations. In addition, choosing images that are more complex to train, in the sense that there is not a clear portion of the image the model can attribute as “positive” or “negative” can allow it to be more versatile and able to take on challenging data sets in the future.

Section 7: Acknowledgements

A special thanks to Imani Musembi, a research mentor who guided me through the intricacies of this project and gave essential feedback to improve this paper from start to finish. To the NIH and clinicians who provided critical datasets used in the entirety of this project, thank you for the hard work and dedication to your craft.

Works Cited

Arishi, Wjdan A., et al. "Techniques for the Detection of Sickle Cell Disease: A Review." *Micromachines*, vol. 12, no. 5, 5 May 2021, p. 519, www.ncbi.nlm.nih.gov/pmc/articles/PMC8148117/, https://doi.org/10.3390/mi12050519. Accessed 7 Sept. 2025.

Brandow, A. M., and R. I. Liem. "Advances in the Diagnosis and Treatment of Sickle Cell Disease." *Journal of Hematology & Oncology*, vol. 15, no. 1, 3 Mar. 2022, jhoonline.biomedcentral.com/articles/10.1186/s13045-022-01237-z, https://doi.org/10.1186/s13045-022-01237-z. Accessed 7 Sept. 2025.

"Cancer Imaging Basics." *Cancer.gov*, 13 Mar. 2025, dctd.cancer.gov/research/research-areas/imaging/basics. Accessed 7 Sept. 2025.

Cafasso, Jacquelyn. "How to Manage a Sickle Cell Crisis." *Healthline*, Healthline Media, 29 Aug. 2018, www.healthline.com/health/sickle-cell-crisis-management.

Chibuta, Samson, and Aybar C. Acar. "Real-Time Malaria Parasite Screening in Thick Blood Smears for Low-Resource Setting." *Journal of Digital Imaging*, vol. 33, no. 3, 23 Jan. 2020, pp. 763–775, https://doi.org/10.1007/s10278-019-00284-2. Accessed 7 Sept. 2025.

Grimberg, Brian T. "Methodology and Application of Flow Cytometry for Investigation of Human Malaria Parasites." *Journal of Immunological Methods*, vol. 367, no. 1-2, 31 Mar. 2011, pp. 1–16, www.ncbi.nlm.nih.gov/pmc/articles/PMC3071436/, https://doi.org/10.1016/j.jim.2011.01.015. Accessed 7 Sept. 2025.

Halenur Sazak, and Muhammed Kotan. "Automated Blood Cell Detection and Classification in Microscopic Images Using YOLOv11 and Optimized Weights." *Diagnostics*, vol. 15, no. 1, 25 Dec. 2024, pp. 22–22, https://doi.org/10.3390/diagnostics15010022.

Hoyos, Kenia, and William Hoyos. "Supporting Malaria Diagnosis Using Deep Learning and Data Augmentation." *Diagnostics*, vol. 14, no. 7, 1 Jan. 2024, p. 690, www.mdpi.com/2075-4418/14/7/690, https://doi.org/10.3390/diagnostics14070690. Accessed 7 Sept. 2025.

Kononenko, Igor. “Machine Learning for Medical Diagnosis: History, State of the Art and Perspective.” *Artificial Intelligence in Medicine*, vol. 23, no. 1, Aug. 2001, pp. 89–109, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.184&rep=rep1&type=pdf, https://doi.org/10.1016/s0933-3657(01)00077-x. Accessed 7 Sept. 2025.

Lei, Yaguo, et al. “Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap.” *Mechanical Systems and Signal Processing*, vol. 138, Apr. 2020, p. 106587, https://doi.org/10.1016/j.ymssp.2019.106587. Accessed 7 Sept. 2025.

Li, Weidong, et al. “Problems Associated with the Diagnosis of Imported Malaria in Anhui Province, China.” *The American Journal of Tropical Medicine and Hygiene*, vol. 102, no. 1, 8 Jan. 2020, pp. 142–146, https://doi.org/10.4269/ajtmh.19-0471. Accessed 7 Sept. 2025.

Osei, Miriam A, and Patrick T McGann. “Sickle Cell Disease: Time to Act on the Most Neglected Global Health Problem.” *The Lancet Haematology*, 1 June 2023, https://doi.org/10.1016/s2352-3026(23)00169-2. Accessed 7 Sept. 2025.

Siłka, Wojciech, et al. “Malaria Detection Using Advanced Deep Learning Architecture.” *Sensors*, vol. 23, no. 3, 29 Jan. 2023, p. 1501, https://doi.org/10.3390/s23031501.

Thakur, Gopal Kumar, et al. “Deep Learning Approaches for Medical Image Analysis and Diagnosis.” *Cureus*, vol. 16, no. 5, 2 May 2024, www.cureus.com/articles/245280-deep-learning-approaches-for-medical-image-analysis-and-diagnosis#, https://doi.org/10.7759/cureus.59507. Accessed 7 Sept. 2025.

Zimmerman, Peter A., and Rosalind E. Howes. “Malaria Diagnosis for Malaria Elimination.” *Current Opinion in Infectious Diseases*, vol. 28, no. 5, Oct. 2015, pp. 446–454, https://doi.org/10.1097/qco.0000000000000191.