1. Introduction

The use of Artificial intelligence has been increasing among numerous fields, including healthcare. A significant application is the use of AI in medical image classification, which offers an alternative to traditional diagnostics. With models that can be trained to current industry accuracy, images can be classified quicker with more accessibility. In the case of malaria, AI models can be trained to identify infected blood cells in microscopy images, rapidly and accurately, without relying on the availability of few specialists.
…*more on the use of AI in image classification*…

In this project, we utilize publicly available datasets containing both parasitized and infected cell images. The primary dataset comes from the National Institute of Health (NIH) and has been made accessible through Kaggle under the title *Malaria Cell Images Dataset.* It consists of 27,558 labeled images divided into two equal classes: 13,779 images of parasitized cells and 13,779 images of uninfected cells. Each image is a microscopic snapshot of a thin blood smear, stained to highlight the presence of Plasmodium parasites, which are the organisms responsible for causing malaria.

The images are categorized into two folders: “Parasitized,” which contains cells visibly infected with malaria parasites, and “Uninfected,” which contains hea;thy red blood cells. Each image is in RGB color format with a resolution of 3 channels (typically 130x130 pixels), and all are stored in JPEG format. The images were collected using a standard light microscope and then digitized and curated by medical professionals.

The dataset is particularly valuable because it is both balanced and pre-labeled, making it ideal for supervised machine learning tasks. It reflects real-world diagnostic scenarios where visual cues, such as color variations, texture changes and distinct spotting, serve as important indicators for infection. By using this dataset, we ensure that our model is trained on clinically relevant and diverse examples, enhancing its potential applicability in automated malaria detection systems.

…*Briefly describe Data Sets and sizes and where they come from and such*…

For this project, our team intends to analyze these datasets through exploratory data analysis, preprocess the data for model training, and build and evaluate a convolutional

neural network capable of accurately classifying malaria infected cells. This project demonstrates the potential of AI to aid in timely, cost-effective malaria diagnosis, especially in resource limited settings. This paper will proceed as follows. Section II will consist of relevant medical background on Malaria, its treatment, and the possible role in AI. In Section III, we conduct exploratory data analysis on the data sets. In Section IV, we will provide details of the AI model we choose to employ. In Section V, we will discuss the implementation of our neural network. Finally, in Section VI, we will review our results and conclude…*Explain what we are going to do with it*…

2. Medical Background

Introduction to Origin

Malaria is a life-threatening infectious disease caused by parasitic protozoans of the genus *Plasmodium*, which are transmitted to humans primarily through the bites of infected female *Anopheles* mosquitoes. The main species responsible for human malaria are *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale*, and *Plasmodium falciparum*.

Symptoms and Effects (Symptoms of Malaria - CDC)

According to the CDC, symptoms of malaria can include fever, flu-like illness, chills, headache, muscle aches, fatigue, nausea, vomiting, and diarrhea. These symptoms can range from very mild to extremely severe, and in some cases, can even lead to death. Individuals may begin to feel ill as early as one week after infection, though symptoms can also take a year or more to appear. As the disease progresses, it can cause anemia (a low red blood cell count) and jaundice (a yellowing of the skin and eyes). If left untreated, malaria can lead to life-threatening complications such as kidney failure, seizures, confusion, coma, and in severe cases, death.

Global Impact (Malaria - an overview)

Malaria remains a major global health crisis, with staggering case numbers and widespread impact. In 2004 alone, between 350 and 500 million cases were reported. More recently, the CDC estimated 249 million malaria cases in 2022, resulting in approximately 608,000 deaths. As one of the most severe public health challenges worldwide, malaria threatens over half of the global population. Sub-Saharan Africa bears the greatest burden,

accounting for the majority of cases and losing more than $12 billion annually due to the disease. Alarmingly, about 25% of deaths among children under five in the region are attributed to malaria. Given the scale and severity of the problem, there is an urgent need to develop new, effective methods for early and accurate malaria diagnosis—particularly in low-resource settings where the disease is most prevalent. Since malaria is both preventable and treatable, improving detection methods could save countless lives, especially among vulnerable populations.

2.1 Diagnosis and Imaging (Malaria Diagnostic Tests | Malaria | CDC)

Microscopic Examination

There are currently several different methods to diagnose malaria, with microscopic examination being the most common and widely accepted. A blood sample taken from the patient is prepared on a microscope slide as either a thick or thin smear, treated with a specialized stain, and examined under a high powered microscope. This process allows for the identification of malaria parasites and, when possible, the distinction between different species. Malaria microscopy is a well established diagnostic technique that relies on basic, widely accessible equipment and materials. It can deliver important clinical information within a few hours of blood collection. When used together, thick and thin smears provide all three essential pieces of information: confirmation of infection, species identification, and parasite density. However, the accuracy of these results depends on the expertise and consistency of the laboratory performing the test. While a negative blood smear generally makes malaria less likely, it does not entirely rule out the disease. Additionally, accurate diagnosis requires the availability of experienced microscopists and timely communication of results to healthcare providers. This is not realistic or ideal for populations experiencing a higher density of the disease.

Rapid Diagnostic Test (RDT)

Rapid Diagnostic Tests (RDTs) offer an alternative method for quickly identifying malaria infections by detecting specific antigens in a patient's blood. To perform the test, a blood sample is applied to the sample pad on the test card along with designated reagents. Within 15 minutes, the appearance of specific indicator bands reveals whether the patient is infected. Because microscopy is not always available in all clinical settings, RDTs can serve as a useful initial tool for rapid diagnosis. However, RDTs are much less accurate than

microscopy and cannot be reliable. So, their use does not replace the need for microscopy. RDTs may fail to detect infections with low parasite densities and have limited sensitivity for the less common species. As a result, all negative RDT results should be confirmed through a follow-up microscopy to rule out infection.

Molecular Detection (PCR)

Another way to determine if a patient is infected with Malaria is through molecular detection. Parasite nucleic acids are detected using polymerase chain reaction (PCR). However, PCR results are not timely and are often not available quickly enough to actually diagnose a malaria infection. Instead, it is more useful for confirming the specific species of malarial parasite after a diagnosis has already been established through microscopy or RDT.

2.2 Treatment and Control (The Treatment of Malaria | New England Journal of Medicine)

Treatment of malaria depends on the infecting species and severity of disease along with the patient's age, immunity, susceptibility to antimalarial drugs, and the cost and availability of treatment.

Treatment Depending on Species

Benign malaria caused by *Plasmodium vivax*, *P. malariae*, and *P. ovale* is typically treated with the drug chloroquine. Although generally well tolerated, chloroquine may cause side effects such as pruritus (itchy skin), nausea, dysphoria, and, in rare cases, neuropathy. For infections involving *P. vivax* and *P. ovale*, a 2-week course of primaquine is administered alongside chloroquine to eradicate hypnozoites—dormant forms of the parasite that persist in the liver. Side effects of primaquine can include nausea, abdominal pain, oxidant hemolysis with methemoglobinemia, anemia, and, in some instances, hemoglobinuria.

Treatment of *P. falciparum* malaria depends heavily on the parasite's local drug sensitivity. In regions such as North Africa, Central America, Haiti, and the Middle East, chloroquine remains the preferred treatment. In areas with low-grade chloroquine resistance, amodiaquine serves as a more effective alternative. In most of Africa, and parts of South America and Asia, a combination of sulfonamide and pyrimethamine is commonly

used. In regions where resistance to antimalarial drugs is high, treatment shifts to medications such as mefloquine, halofantrine, or quinine combined with tetracycline.

Treatment of Severe Malaria

The earlier the diagnosis, the less complicated the treatment. Fatalities due to *Plasmodium vivax, P. malariae,* and *P. ovale* are rare; however, infections with *P. falciparum* can progress to multisystem disease. Clinical manifestations of severe malaria vary by age and can range from hypoglycemia and severe anemia to cerebral malaria and respiratory arrest.

Doctors pay special attention to patient hydration, carefully avoiding both fluid overload and underhydration. Failure to control hydration levels can lead to hypotension and shock, accelerating the onset of acute renal failure. In cases of vital-organ dysfunction, dialysis or hemofiltration is required.

Following rehydration, patients receive glucose maintenance infusions with constant monitoring of their blood glucose levels. Unconscious cerebral malaria patients are positioned on their side to reduce the risk of seizures. If seizures occur, intravenous benzodiazepines are administered. In other extreme cases, blood transfusions may be necessary.

It is not uncommon for patients with severe malaria to develop secondary bacterial infections such as pneumonia, urinary tract infections, septicemia, and systemic salmonella. In such cases, broad-spectrum antimicrobial agents are initiated.

Throughout treatment, vital signs are monitored as frequently as possible. These include the patient's coma score, urine output, blood glucose level, lactate level, arterial pH, blood gas levels, and parasite count. If the parasite count has not decreased by at least 75 percent within 48 hours of starting treatment, a different antimalarial agent is administered.

Importance of AI

Early detection of malaria is essential to prevent serious health complications and reduce mortality rates. Artificial intelligence offers a powerful, cost-effective, and rapid tool for early diagnosis, particularly in resource-limited settings. Beyond simply detecting the presence of malaria, accurately identifying the specific species is critical, as treatment protocols vary depending on the species involved. Misdiagnosis can lead to ineffective

treatment and increased risk of severe outcomes. By leveraging AI to quickly and accurately distinguish between malaria species, we can greatly improve patient outcomes.

2.3 The Role of AI in Malaria Detection (Nationwide real-world implementation of AI for cancer detection in population-based mammography screening | Nature Medicine)

Traditional methods of malaria detection require significant expertise, time, and manpower. However, AI-driven diagnostic tools have the potential to overcome many of these hurdles by providing fast, scalable, and accurate analysis of blood samples. Artificial Intelligence is already being used in medical diagnostics. For instance, AI algorithms are being employed to analyze images and aid in the detection and classification of various cancers with high accuracy. These AI systems can help identify subtle patterns in imaging data that may be overlooked by the human eye. This has been proven to enhance diagnostic precision. Applying these same techniques and principles to create an AI to detect the presence of Malaria could help aid in many of the current limitations of the traditional diagnosis techniques. AI systems can process and analyze images rapidly, providing quicker diagnostic results, achieve higher levels of accuracy, potentially reducing the proportion of false negatives and positives, and can potentially aid in the classification of the parasite, eliminating the need for PCR which is time consuming and expensive.

3. Exploratory Data Analysis (Diego, Deethya, & Matthew)

3.1 Data Source

The dataset for this project was sourced from Kaggle's Malaria Cell Images Dataset, specifically available through the Kaggle notebook directory: Malaria Cells Classification. This dataset consists of microscopic images of blood cells labeled as either "Parasitized" or "Uninfected." We accessed and used two distinct subdirectories: Data Sources/Malaria Cell images/Parasitized and Data Sources/Malaria Cell images/Uninfected. These images served as the foundational data for our classification task. Each image captures a single red blood cell, and the presence or absence of parasitic infection provides the label needed for supervised learning. The goal of our project was to extract key visual features from these images and use them to distinguish between infected and uninfected cells.

By collecting this data, we established a labeled dataset suitable for feature extraction, analysis, and training machine learning models. The images provided a controlled and consistent input space in terms of format, resolution, and biological context, which was critical for applying standardized preprocessing and analysis techniques. The visual differences—such as spot formation, color shifts, or structural texture anomalies—became the main focus of our data cleaning and preparation strategy.

Data Cleaning Strategy

Before applying any classification model, it was essential to preprocess and clean the image data to isolate the relevant features for analysis. Visual inspection revealed that infected cells exhibited distinguishable patterns—often in the form of purple "spots"—that were not present in uninfected cells. These observations guided the selection of the features we aimed to extract: texture features, color features, and pixel intensity.

Texture features were used to capture spatial patterns within the images. These features quantify the spatial arrangement and relationships of pixel values across a region or the entire image. By analyzing texture, we were able to detect subtle structural differences such as clustering, granularity, or irregular arrangements that might indicate parasitic infection. Texture analysis is particularly valuable for biomedical imaging tasks, as the shape and distribution of anomalies can serve as strong diagnostic indicators.

Color features, on the other hand, provided spectral information that is directly derived from the image's RGB channels. Since the images are colored and the staining used in microscopy highlights parasitic regions in distinguishable hues (often purples or reds), color becomes a key differentiator. We extracted RGB values from each pixel, allowing us to quantify differences in color intensity and distribution across infected and uninfected cells. Each pixel's red, green, and blue values—typically stored in 8-bit resolution—offered a rich source of information for classification.

Pixel intensity, another critical feature, refers to the brightness level at each pixel. For grayscale images, this is a single value; for color images, it refers to the combined intensity across the RGB channels. Pixel intensity is one of the most fundamental and informative properties in image classification. By analyzing the intensity values of each pixel, we were

able to quantify contrast and highlight regions of interest, such as infected spots. These values were crucial for identifying and differentiating infected areas based on their brightness and darkness compared to the surrounding cellular structure.

3.1 Code Implementation for Cleaning and Feature Extraction

To implement our data cleaning pipeline, we used a Python environment accessible through a terminal. We selected Python 3.8.2 as our working environment and utilized core libraries such as OpenCV and NumPy, which are well-suited for image analysis. Since OpenCV was not pre-installed in our environment, we first installed it using the pip package manager with the command: python3 -m pip install --user opencv-python. This installed the necessary packages locally within the user directory, allowing us to run the scripts without elevated permissions.

With the environment set up, we proceeded to write and execute Python scripts directly from the terminal using the nano text editor. Three primary scripts were developed for the cleaning and preprocessing phase. The first script, pixelintensity.py, was responsible for analyzing and exporting pixel intensity values from each image into a structured CSV format. The second script, texturefeatures.py, extracted spatial texture metrics and saved them for further analysis. Finally, dataclean.py was designed to handle color feature extraction by iterating through image files and capturing RGB channel statistics.

While executing these scripts, we encountered a common coding error related to incorrect method usage (AttributeError: 'str' object has no attribute 'rglob'), which we quickly addressed by modifying the way files were accessed and read. After resolving this, all scripts ran successfully and produced corresponding CSV outputs. These structured data files contained the cleaned and quantified features from our original image set and became the input for our downstream classification algorithms.

References

Possibly useful sources:

- Malaria – an overview
- The Treatment Of Malaria
- World Malaria Report 2005
- Malaria Detection using Deep Learning
- Malaria Detection Using Advanced Deep Learning Architecture
- COMPARISON OF FIVE METHODS OF MALARIA DETECTION…
- How Malaria Has Affected the Human Genome and What Human Genetics...
- Nationwide real-world implementation of AI for cancer detection in population-based mammography screening | Nature Medicine
- Deep learning-based classification of lymphedema and other lower limb edema diseases using clinical images | Scientific Reports
- Symptoms of Malaria - CDC
- Malaria Diagnostic Tests | Malaria | CDC