Executive Summary

Identifying Malaria in Patient Blood Samples Using Convolutional Neural Networks

Research Question

Can convolutional neural networks (CNNs) be used to identify malaria-infected red blood cells in patient blood samples?

Context

Malaria kills hundreds of thousands of people every year and is challenging to diagnose in resource-limited settings. The standard method is microscopic examination of a Giemsa-stained blood sample by a trained microbiologist. While effective, it is impractical where resources, particularly trained personnel, are scarce. This results in delayed or missed diagnoses, increasing mortality. Developing a faster, accurate diagnostic method that reduces reliance on trained professionals could improve treatment access and save lives. This capstone project proposes the use of CNNs to accurately process blood work images to diagnose people with malaria as an alternative diagnostic tool.

Hypotheses

The null hypothesis states that the convolutional neural network cannot predict cells infected with malaria with statistical significance, achieving at least 90% accuracy. Conversely, the alternative hypothesis states that the convolutional neural network can predict cells infected with malaria with statistical significance, achieving at least 90% accuracy.

Data Analysis

The images for this project were downloaded from Kaggle and organized into folders for healthy and malaria-infected cells. They were split into training, validation, and testing sets with 17,879, 4,469, and 5,512 images, respectively. The images were scaled to 192x192 pixels using `TensorFlow`'s `ImageDataGenerator()` function. This approach maintained image details and prepared the data for use in TensorFlow, resulting in a properly divided and scaled dataset for training, validating, and testing the convolutional neural network (CNN).

The initial model was a simple neural network with only input and output layers, achieving a validation accuracy of 0.512. Adding a convolutional layer with 16 filters and ReLU activation improved validation accuracy to 0.724 but led to overfitting. Introducing a dropout layer reduced overfitting and increased validation accuracy. Adding a hidden dense layer significantly increased model size, which was managed by incorporating a max pooling layer, achieving 0.897 validation accuracy. Further refining with an additional convolutional layer and a final dense layer, the CNN achieved a final validation accuracy of 0.934 and training accuracy of 0.964.

Findings

After using the final model to make predictions on the testing images, the p-value obtained from the chi-square test on the confusion matrix of the predictions is approximately zero and the overall accuracy of the model on the testing images is 94.4%. Therefore, the null hypothesis is rejected. This result indicates that the CNN can accurately classify blood cell images as either healthy or malaria-infected.

Limitations of Techniques and Tools

A limitation of this project is the requirement for the technology to host the CNN at testing sites, which may not always be feasible due to the lack of necessary infrastructure, such as reliable electricity, in remote and under-resourced areas. Addressing this limitation will require careful consideration of each potential testing location and its unique challenges. Possible solutions could include creating portable, self-sufficient diagnostic devices that can operate independently of existing infrastructure.

Proposed Actions

The recommended course of action is to partner with humanitarian and medical aid programs to assess how and where this CNN can address needs and relieve strain on the medical system in the areas they serve. These collaborations can identify the most beneficial deployment areas and integrate the CNN into existing healthcare frameworks, ensuring effective and sustainable implementation to improve malaria diagnosis in resource-limited settings.

Expected Benefits

The implications of this finding are significant for malaria-prevalent regions with limited resources and microbiologists. Automating the diagnostic process with this CNN model offers a reliable alternative to traditional microscopic examination that relies on trained personnel, accelerating diagnostics, improving access and speed to treatment, and potentially reducing malaria-related deaths. This success also highlights machine learning's potential in other global health initiatives.