CSE/ECE 343 : Machine Learning Project Proposal
Title : Detecting Malaria using Machine Learning

Srishti Singh
srishti20409@iiitd.ac.in

Shreya Bhatia
shreya20542@iiitd.ac.in

Madhava Krishna
madhava20217@iiitd.ac.in

Harshit Goyal
harshit20203@iiitd.ac.in

1. Motivation

Malaria is a life-threatening disease affecting many people worldwide, spread by infected Anopheles mosquito bites. Earlier studies have shown that the degree of agreement between physicians on the acuteness of the disease in a given patient's sample is very low. Preliminary detection aided by computer systems can be of utmost importance for faster and reliable diagnosis. We aim to create a classifier for parasitized and non-parasitized cells to aid medical professionals in this venture.

2. Related Work

- Pan, et al. (2018) created a model based on deep CNN architectures. They were able to obtain accuracies of over 90% on the training and validation samples using data augmentation.
- Raihan and Nahid (2021) created a model based on boosted trees with feature engineering and determined feature importance using Shapely Additive Explanations (SHAP).
- Fulhad et al. (2020) implemented a CNN based model with accuracy over 99% while being computationally efficient.

3. Timeline

A Tentative 12 week timeline:

Week 1-2	Exploring the domain
Week 3	Data Collection
Week 4	Pre-processing and Data Visualization
Week 5	Feature Extraction
Week 6	Logistic Regression, Support Vector Machines
Week 7	Decision Trees, Random Forest
Week 8	Analysis & performance of models
Week 9-10	Hyperparameter Tuning, Check for Overfitting, Underfitting
Week 11	Report Writing.
Week 12	Buffer

4. Individual Tasks

The individual task have been distributed in following manner -

Data Collection	Srishti, Madhava, Harshit, Shreya
Data Visualisation , Pre Processing	Shreya . Madhava
Logistic Regression	Shreya, Harshit
Decision tress & Random forest	Madhava, Srishti
Analysis and Performance of model	Srishti, Madhava, Harshit, Shreya
Feature extraction	Harshit, Shreya
CNN	Madhava, Srishti
Hyperparameter Tuning	Srishti, Shreya, Harshit
Check for Overfitting, underfitting	Srishti, Shreya, Harshit
Report Writing	Srishti, Madhava, Harshit, Shreya

5. Final Outcome

Automation of the diagnosis process will guarantee accurate diagnosis and, as a result, holds the possibility of providing dependable healthcare to places with limited resources. We aim to implement various algorithms for classification while attempting to find optimal parameters for optimising training time, computational complexity and performance. We will attempt transformations and feature engineering and extraction on the dataset. We are going to apply various machine learning models such as SVM, Logistic regression, decision trees, Random Forest, and compare the performances of all the models. We intend to convert the images to grayscale and see the change in behavior of the model.