Malaria Forecasting Using Deep Learning and Epidemiological Data

A Multi-Phase Approach Combining Image Classification and Outbreak Prediction

Cheikhani (C20114)

Introduction

Malaria remains one of the most severe infectious diseases in tropical and subtropical regions worldwide. Rapid and accurate diagnosis is essential for timely treatment and control. This project integrates deep learning techniques for automated malaria detection from microscopic images, with a planned expansion to forecast malaria outbreaks using epidemiological and climatic data.

Methods

1. Image Classification Using CNN (Phase 1: Initial Prototype) A subset of 1,000 microscopic red blood cell images was used, equally labeled as "infected" (500) and "uninfected" (500). Images were resized to 128×128 pixels, and data augmentation techniques such as random rotations and flipping were applied to improve model robustness.

A simple CNN architecture composed of Conv2D, MaxPooling2D, and Dense layers was built. The model was compiled with the Adam optimizer and trained using Binary Crossentropy loss.

2. Planned Enhancement – Transfer Learning (Phase 2) The next phase will leverage a larger dataset exceeding 25,000 images and apply transfer learning with pretrained models such as MobileNetV2 (lightweight and fast) and ResNet50 (deep and accurate). These models will be fine-tuned for malaria detection, and interpretability methods like Grad-CAM will be used to visualize decision areas.

Results (Phase 1 – CNN on 1,000 Images)

- Accuracy: 92%
- Precision: 91%
- Recall: 93%
- F1-Score: 92%

The initial prototype successfully distinguished infected from uninfected cells, demonstrating the potential of CNNs in malaria diagnosis.

Future Work

3. Malaria Outbreak Forecasting Module To increase the societal impact, a forecasting module will be developed to predict regional malaria outbreaks using epidemiological and climatic data. *However, this phase will commence only if a suitable dataset containing relevant epidemiological and climate data can be obtained.*

3.1 Data Sources Datasets will be sourced from Kaggle (malaria and climate data), WHO and OpenAFRICA (regional and global malaria statistics). Key features will include temperature, rainfall, humidity, historical case counts, and geographic location.

3.2 Candidate Models

- XGBoost and Random Forest for tabular predictive modeling
- Prophet and ARIMA for time series forecasting
- LSTM networks for deep learning on sequential data

3.3 Geospatial Visualization Risk heatmaps and outbreak zones will be visualized using GIS libraries such as geopandas, folium, or Plotly for interactive exploration.

4. Deployment Strategy A user-friendly Streamlit application will integrate all components, featuring:

- Upload and classify blood cell images with confidence scores
- Grad-CAM heatmaps to interpret CNN predictions
- Dashboard for climate-based malaria outbreak forecasting
- Interactive maps showing predicted spread in real time

Conclusion

The initial CNN-based malaria detection model trained on a small dataset yielded promising results. Upcoming phases aim to scale the classifier with transfer learning and construct an epidemiological forecasting system conditional on acquiring suitable datasets. This integrated approach has the potential to enhance malaria control efforts, especially in resource-limited settings.