CSE 499A: Project Work | Section: 15 | Group: 07 | Members: Amit Chandra Das (2014242042), Tanushree Das (2212225042), Ashraful Islam (2212669042)

Project Progress Report: Arsenic Skin Detection using Machine Learning

Overall Summary: Today, the team successfully completed the entire pipeline for building, training, and evaluating our baseline Convolutional Neural Network (CNN) model. The project moved from raw image files to a fully functional and evaluated classifier, achieving a strong initial test accuracy of **85.44%**. The work was partitioned into three main phases as follows:

Tanushree Das: Data Processing and Final Preparation

This phase focused on converting the raw image dataset into a clean, standardized, and machine-learning-ready format.

1. **Image-to-Array Conversion:** All 8,892 images (both 'infected' and 'not_infected') were loaded from disk. Each image was systematically processed by:
    - **Resizing:** Standardizing all images to 128x128 pixels.
    - **Color Standardization:** Converting all images to the 'RGB' (3-channel) format to ensure uniformity.
    - **Numerical Conversion:** Transforming each 128x128 image into a NumPy array with the shape (128, 128, 3).
2. **Dataset Splitting:** The complete numerical dataset (data) and corresponding labels were split into an 80% training set (for teaching the model) and a 20% testing set (for evaluation).
3. **Normalization:** All pixel values in both the training and testing sets were scaled from the [0-255] range to the [0.0-1.0] range by dividing by 255.0. This is a critical step for optimizing model training.

All data is fully processed and ready for modeling.

Ashraful Islam: CNN Model Architecture and Compilation

This phase involved designing and building the "brain" of the operation the neural network architecture.

1. **Model Instantiation:** A Sequential model was built using the Keras library.
2. **Feature Extraction Layers:** The core of the CNN was constructed using two blocks of Conv2D (convolutional) and MaxPooling2D layers. These layers are responsible for automatically finding and learning patterns (like texture and edges) in the skin images.
3. **Classification Layers:**
    - A Flatten layer was used to transition the 2D image data into a 1D format.

- A Dense (fully-connected) layer was added to perform high-level classification.
- A Dropout layer was included to help prevent overfitting.
- The final Dense (output) layer uses softmax activation to produce the final probability for 'infected' or 'not_infected'.

4. **Model Compilation:** The model was compiled with the 'adam' optimizer and the 'sparse_categorical_crossentropy' loss function, which are standard and effective choices for this type of classification problem.

The model architecture is defined and ready for training.

Amit Chandra Das: Model Training, Evaluation, and Results Analysis

This phase focused on training the compiled model and rigorously evaluating its performance on unseen data.

1. **Model Training:** The model.fit() function was called, training the CNN on the prepared training data for 10 full epochs. The model learned to differentiate between the two classes by adjusting its internal parameters over these 10 cycles.
2. **Performance Evaluation:** After training, the model's performance was measured against the 20% test set, which it had never seen before.
3. **Key Result:** The model achieved a final **test accuracy of 85.44%**. This indicates it correctly classified the images in the test set over 85% of the time.
4. **Results Analysis:** To understand the model's behavior, its training and validation accuracy/loss were plotted. A final classification report was also generated, providing detailed metrics (like precision and recall) for both the 'infected' and 'not_infected' classes, confirming the model is well-balanced.