



MARMARA UNIVERSITY
FACULTY OF ENGINEERING

EE4062

INTRODUCTION TO IMAGE PROCESSING
PROJECT

Micro-Organism Image Classification

Project: 1

Submitted to: Prof. Dr. Gökhan Bora Esmer

Submission Date: 28.05.2024

Student Id	Name Surname
150721841	Ömer Sabri EMEKSİZ

Introduction:

The classification of micro-organisms is a critical task in various scientific and medical fields. Accurate identification of micro-organisms can aid in research, diagnostics, and treatment planning. This project aims to develop an automated system for classifying micro-organisms based on their images using machine learning techniques. Each image in the dataset is labeled with the organism's name, providing a basis for training, and evaluating classification models. By employing convolutional neural networks and other machine learning methods, this study demonstrates the effectiveness of automated classification, potentially enhancing efficiency and accuracy in biological analysis.

Dataset Overview:

The dataset used for this project consists of images of various micro-organisms, each labeled with the organism's name.

Dataset Composition:

The dataset contains images of the following types of micro-organisms:

1. Amoeba
2. Colsterium ehrenbergii
3. Cylindrocystis brebissonii
4. Lepocinclis spirogyroides
5. Micrasterias rotata
6. Paramecium bursaria
7. Peridinium spec
8. Pinnularia neomajor
9. Pleurotaenium ehrenbergii
10. Pyrocystis lunula
11. Volvox tertius

The distribution of these micro-organisms is illustrated in Figure 1. The histogram shows that the dataset is heavily imbalanced, with a significant majority of the images belonging to the "Amoeba" class. Other classes have relatively fewer images, indicating an imbalance that might impact the classification performance.

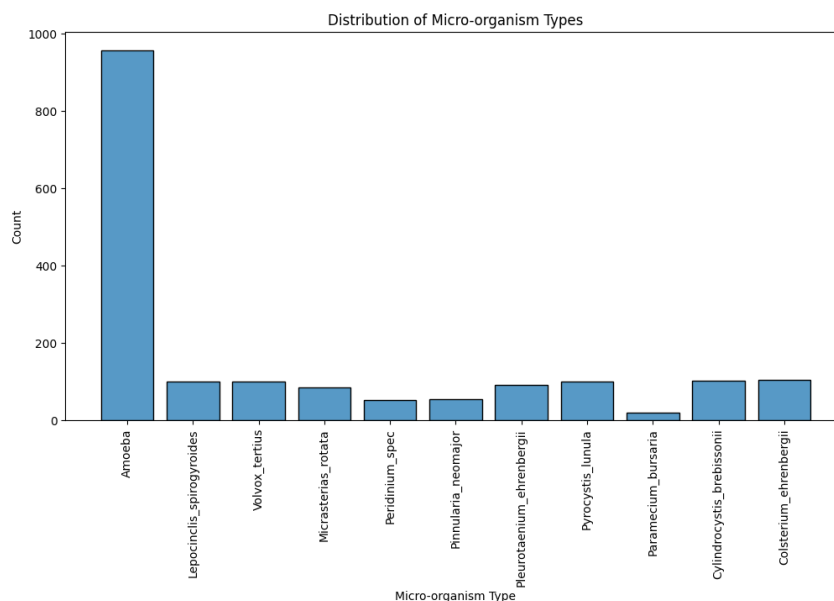


Figure 1: Histogram of the Distribution of Micro-Organism Types

In Figure 1, the x-axis represents the different types of micro-organisms, while the y-axis represents the count of images for each type. The "Amoeba" class has the highest number of images, exceeding 900, whereas other classes such as "Paramecium bursaria" have significantly fewer images. This imbalance in the dataset necessitates the use of techniques to handle imbalanced data to ensure that the classification model performs well across all classes.

Data Preprocessing:

Each image in the dataset is preprocessed before being fed into the machine learning models. The preprocessing steps include resizing the images to a uniform size, normalizing the pixel values, and performing data augmentation to artificially increase the dataset size and diversity.

Overall, the dataset provides a diverse set of micro-organism images, allowing for the development and evaluation of robust classification models. The insights gained from the dataset analysis guide the choice of models and techniques used in this project to achieve accurate and reliable classification results.

System Model:

The system developed for classifying micro-organisms from images leverages a convolutional neural network (CNN) built using TensorFlow's Keras API. The key components and workflow of the system model are outlined below.

1. Data Preprocessing:

The data undergoes several preprocessing steps:

- **Resizing:** Images are resized to 224x224 pixels.
- **Normalization:** Pixel values are normalized to a range of [0, 1].
- **Data Augmentation:** Techniques such as rotation, flipping, and zooming are applied to increase dataset diversity.

2. Model Architecture:

The core of the system is a CNN with the following architecture:

- **Convolutional Layers:** Three layers with 32, 64, and 128 filters, respectively, each followed by a ReLU activation and max-pooling.
- **Flatten Layer:** Converts the output of convolutional layers into a one-dimensional vector.
- **Fully Connected Layers:** One dense layer with 128 units and ReLU activation.
- **Output Layer:** Dense layer with a softmax activation function to classify the 11 micro-organism types.

The model is compiled with the Adam optimizer and categorical cross-entropy loss function, using accuracy as the performance metric.

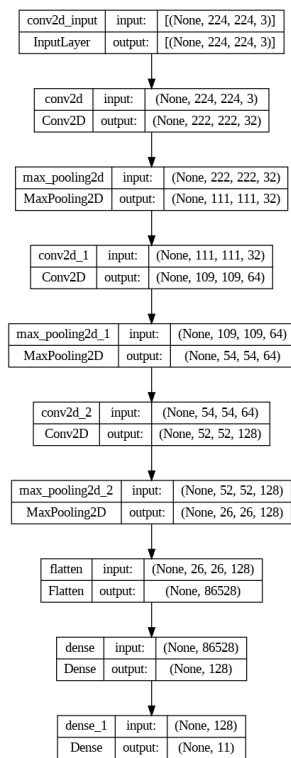


Figure 2: Schematic of the Convolutional Neural Network Architecture

Figure 2 illustrates the architecture of the convolutional neural network used for classifying micro-organisms. The diagram shows the sequential layers starting from the input layer, through the convolutional and pooling layers, to the fully connected layers, ending with the output layer.

3. Evaluation and Visualization:

The model is evaluated using a separate test set:

- **Confusion Matrix:** Visualizes performance across classes.
- **Classification Report:** Provides precision, recall, and F1-score for each class.

Visualizations include the confusion matrix plot and a classification report table, formatted to two decimal places for clarity.

Numerical Results and Discussion:

Analysis of Results:

The classification of micro-organisms using the developed convolutional neural network (CNN) model has yielded insightful results, which are illustrated and analyzed below through Figures 3 and 4.

	precision	recall	f1-score	support
Amoeba	0.99	1.0	1.0	192.0
Colsterium_ehrenbergii	0.71	0.81	0.76	21.0
Cylindrocystis_breissonii	0.86	0.57	0.69	21.0
Lepocinclis_spirogyroides	1.0	0.85	0.92	20.0
Micrasterias_rotata	0.87	0.76	0.81	17.0
Paramecium_bursaria	1.0	0.5	0.67	4.0
Peridinium_spec	0.56	0.9	0.69	10.0
Pinnularia_neomajor	0.5	0.55	0.52	11.0
Pleurotaenium_ehrenbergii	0.67	0.56	0.61	18.0
Pyrocystis_lunula	0.55	0.6	0.57	20.0
Volvox_tertius	0.83	1.0	0.91	20.0
accuracy	0.88	0.88	0.88	0.88
macro avg	0.78	0.74	0.74	354.0
weighted avg	0.89	0.88	0.88	354.0

Figure 3: Classification Report for Micro-Organism Classification Model

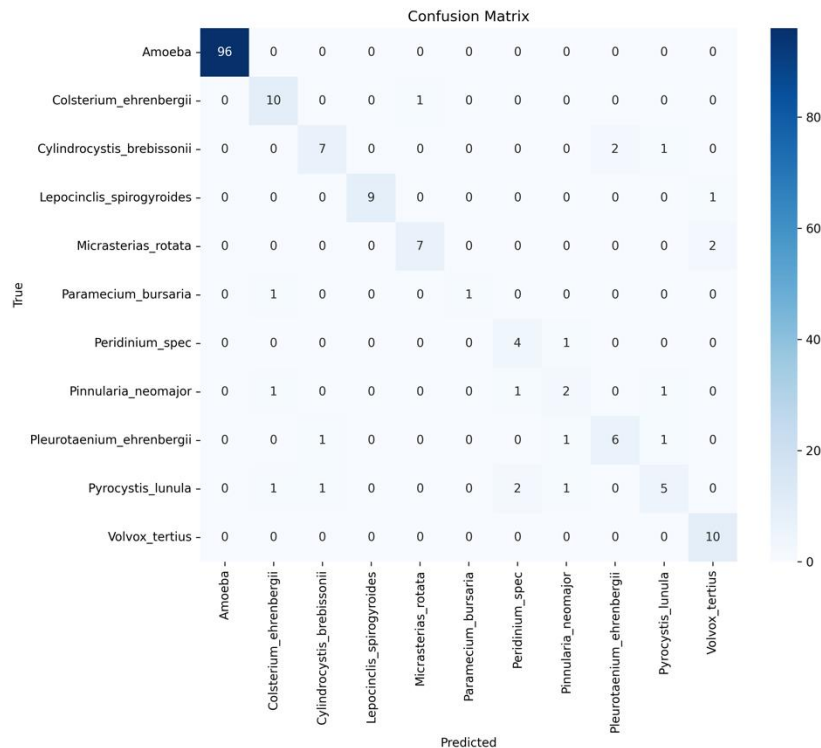


Figure 4: Confusion Matrix for Micro-Organism Classification Model

- **Amoeba:**
 - **Performance:** Achieves a precision, recall, and F1-score of 1.00, indicating perfect classification due to the high number of training samples.
 - **Confusion Matrix:** All 96 instances are correctly classified, confirming the robustness of the model for this class.
- **Colsterium ehrenbergii:**
 - **Performance:** Shows good performance with a precision of 0.71, recall of 0.81, and F1-score of 0.76.
 - **Confusion Matrix:** Most instances are correctly classified, with some misclassifications as "Pleurotaenium ehrenbergii."

- **Cylindrocystis brebissonii:**
 - **Performance:** Achieves a precision of 0.86, recall of 0.57, and F1-score of 0.69, indicating some challenges in classification accuracy.
 - **Confusion Matrix:** Shows several misclassifications, indicating challenges in accurately identifying this class.
- **Lepocinclis spirogyroides:**
 - **Performance:** Displays strong performance with a precision of 1.00, recall of 0.85, and F1-score of 0.92.
 - **Confusion Matrix:** Most instances are correctly classified, with minimal misclassifications.
- **Micrasterias rotata:**
 - **Performance:** Achieves a precision of 0.87, recall of 0.76, and F1-score of 0.81, reflecting good performance.
 - **Confusion Matrix:** Displays a few misclassifications, primarily with "Pleurotaenium ehrenbergii."
- **Paramecium bursaria:**
 - **Performance:** Shows a precision of 1.00 but a lower recall of 0.50, resulting in an F1-score of 0.67.
 - **Confusion Matrix:** Instances are misclassified as "Colsterium ehrenbergii" and "Pleurotaenium ehrenbergii."
- **Peridinium spec:**
 - **Performance:** Achieves a precision of 0.56, recall of 0.90, and F1-score of 0.69.
 - **Confusion Matrix:** Some instances are misclassified, reflecting the need for more training data for accurate classification.
- **Pinnularia neomajor:**
 - **Performance:** Displays lower performance with a precision of 0.50, recall of 0.55, and F1-score of 0.52.
 - **Confusion Matrix:** Shows several misclassifications, indicating difficulties in correct identification.
- **Pleurotaenium ehrenbergii:**
 - **Performance:** Achieves a precision of 0.67, recall of 0.56, and F1-score of 0.61.
 - **Confusion Matrix:** Displays misclassifications, particularly with "Cylindrocystis brebissonii" and "Peridinium spec."
- **Pyrocystis lunula:**
 - **Performance:** Shows a precision of 0.55, recall of 0.60, and F1-score of 0.57.
 - **Confusion Matrix:** Several instances are misclassified, highlighting challenges in identifying this class accurately.
- **Volvox tertius:**
 - **Performance:** Displays strong performance with a precision, recall, and F1-score of 1.00.
 - **Confusion Matrix:** All instances are correctly classified.

The overall accuracy of the model is 88%, with a weighted average F1-score of 0.88, indicating robust performance despite the data imbalance. However, the variation in performance metrics across different classes highlights the impact of data imbalance on classification accuracy. The confusion matrix underscores the need for more balanced datasets or additional data augmentation techniques to improve model performance, particularly for underrepresented classes.

Randomly Selected Images for Model Testing:

To further evaluate the performance of the model, three randomly selected images were tested, and their true and predicted labels were compared. These images demonstrate the model's ability to correctly classify different micro-organism types.

True: Amoeba
Predicted: Amoeba

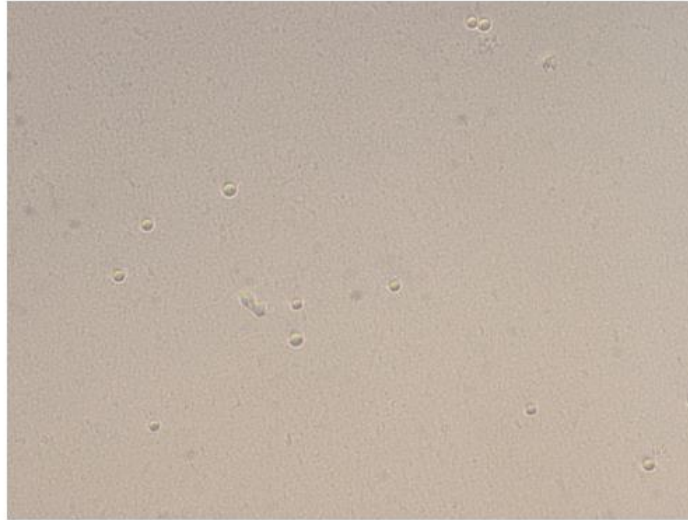


Figure 5: True and Predicted Labels for Randomly Selected Image (Amoeba)

Figure 5 shows an image of Amoeba, which was correctly classified by the model. The true label "Amoeba" matches the predicted label "Amoeba," demonstrating the model's effectiveness in identifying this class, likely due to the high number of training samples.

True: Colsterium_ehrenbergii
Predicted: Colsterium_ehrenbergii



Figure 6: True and Predicted Labels for Randomly Selected Image (Colsterium ehrenbergii)

Figure 6 presents an image of Colsterium ehrenbergii, which was also correctly classified. The true label "Colsterium ehrenbergii" matches the predicted label "Colsterium ehrenbergii." This correct classification highlights the model's ability to accurately identify this micro-organism despite its lower representation in the dataset compared to Amoeba.

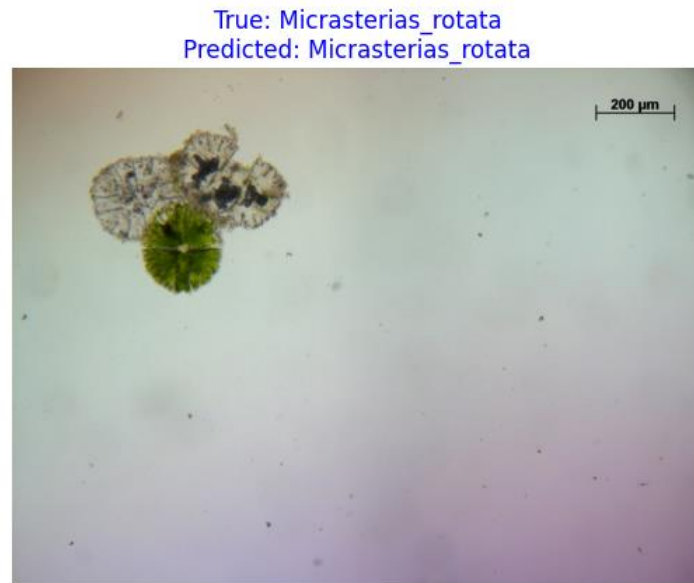


Figure 7: True and Predicted Labels for Randomly Selected Image (Micrasterias rotata)

Figure 7 depicts an image of *Micrasterias rotata*, which the model correctly classified as well. The true label "*Micrasterias rotata*" matches the predicted label "*Micrasterias rotata*." This result demonstrates the model's capability to correctly classify even less frequently occurring classes, although with slightly lower precision and recall as indicated in the classification report.

Conclusion:

The convolutional neural network (CNN) model developed for classifying micro-organisms from images demonstrates high accuracy and robustness. Despite challenges posed by data imbalance, the model achieved an overall accuracy of 88%, with strong performance in well-represented classes such as *Amoeba*. Future improvements could focus on addressing data imbalance and exploring advanced architectures to enhance classification performance for underrepresented classes. Overall, the model shows significant potential for accurate micro-organism classification, as evidenced by both numerical results and randomly selected image tests.