Analysis of Heatmap Images: Supervised and Unsupervised Learning Tasks

TTT23 Biomedical Signal and Image Processing, and Communications

1 Introduction

This report presents two machine learning assignments aimed at analyzing a dataset of 33 images of mice with tumors. The images have been processed to display heatmaps representing the bioluminescence signal intensity resulting from an injection of mammary carcinoma cells fraction enriched with extracellular vesicles. The assignments are designed to predict signal intensity values and extract meaningful features from the images, respectively.

2 Supervised Learning Task

2.1 Objective

Predict the signal intensity values for the head, lung, and abdomen regions in the images using supervised learning techniques.

2.2 Methodology

1. **Data Preparation:** Normalize or standardize the pixel values, ensure correct labeling of intensity values.
2. **Data Augmentation:** Apply random rotations, zooming, flipping, and brightness adjustments.
3. **Model Selection and Training:** Utilize a simpler or pre-trained model, divide the augmented dataset into training and validation sets.
4. **Regularization:** Apply L1, L2 regularization or Dropout to prevent overfitting.

5. **Early Stopping:** Implement early stopping with a patience parameter.
6. **Evaluation:** Utilize metrics like Mean Absolute Error (MAE) or Mean Squared Error (MSE).
7. **Optimization:** Adjust hyperparameters or try different architectures to improve performance.

2.3 Challenges

The primary challenges include overfitting due to the limited dataset and inaccurate estimation of model performance due to limited validation data.

3 Unsupervised Learning Task

3.1 Objective

Extract meaningful features from the images capturing the variations in signal intensity and other tumor characteristics.

3.2 Methodology

1. **Data Preprocessing:** Similar preprocessing steps as in the Supervised Learning Task.
2. **Data Augmentation:** Apply the same data augmentation techniques.
3. **Feature Extraction:** Utilize PCA or Autoencoders, apply regularization if using Autoencoders.
4. **Model Complexity:** Opt for simpler models or configurations.
5. **Evaluation:** Evaluate the usefulness of the extracted features using domain-specific metrics or visualization techniques.
6. **Clustering (Optional):** Employ clustering algorithms like K-means.
7. **Visualization:** Use techniques like t-SNE or UMAP for visualization.

3.3 Challenges

The primary challenges include interpretability of the extracted features and insufficient data leading to less meaningful or less diverse features extraction.

4 Conclusion

The assignments provide a structured approach to analyzing the heatmap images for predicting signal intensity values and extracting meaningful features. Despite the limited dataset size, the outlined methodologies aim to leverage data augmentation and regularization techniques to mitigate overfitting and extract valuable insights from the data.