Machine Learning Project Proposal

Group members: Zongzhu Li, Maliha Aziz

- What problem did you select, and why did you select it?

  To sharpen our machine learning skills and understand real-world data , our group decided to select the topic of detecting brain tumors in Magnetic Resonance Imaging from a Kaggle competition: https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri

  The objective of this project is trying to classify the tumor types after training all the MRI images.

  The reason we select it is because automated classification techniques using Machine Learning(ML) and Artificial Intelligence(AI) has consistently shown higher accuracy than manual classification. Hence, proposing a system performing detection and classification by using Deep Learning Algorithms using Artificial Neural Network (ANN) would be helpful to doctors all around the world.

- What database will you use? Is it large enough to train a machine learning network or different algorithms?

  We would use the dataset provided by Kaggle. This dataset is collected in the form of images through the scans which is the best technique to detect brain tumors. The dataset needs to be read out into matrix data. After that, we will have lots of data which is large enough to train a machine learning network.

- What neural network will you use? Will it be a standard form of the network, or will you have to customize it? What algorithms will you use?.

  The ANN classification algorithm will be used in this project according to the objects. Some of the parameters, like layer number, learning rate, will be modified to make a comparison.

- What software will you use to implement the neural network or different algorithms? Why?

  Python will be adopted here to implement the neural network because it is easy to use and many packages are plug-in-use.

- What reference materials will you use to obtain sufficient background on applying the chosen network or algorithm to the specific problem that you selected?

In this project, EDA, preprocessing, image reading out, model building and evaluation strategy, will be applied, so all the knowledge and reference materials related to the above topic will be referred.

- How will you judge the performance of the network? What metrics will you use?

  After building a model with the 80% randomly selected data from the training dataset, we will apply the model to the 20% left dataset to evaluate the model performance and fitness. The evaluation metrics will be the accuracy score.

- Provide a rough schedule for completing the project.

  June 15 - 16: proposal writing and data preparation;
  June 17 - 20: code for model building and evaluation strategy;
  June 21 - 23: presentation preparation and final report.