

Model Testing and Training

Submitted by:

Group Number 26 [Manikanta Desu(203174002), Mayank Seth(203190001), Vivek Rathore(190100138), Ashara Keyur Ashwin(19D100003), Akash Rajendra Gayakwad(190260004)]

For the project following libraries are mainly used:

1) Pandas: This library is used for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables. This library is also used to loading the datasets from different formats

2) OpenCV: OpenCV is a cross-platform library using which we can develop real time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection.

3) TensorFlow: TensorFlow is a Python library for fast numerical computing created and released by Google. It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of TensorFlow.

4) Keras: Keras is a powerful and easy-to-use free open source Python library for developing and evaluating deep learning models. It wraps the efficient numerical computation libraries Theano and TensorFlow and allows you to define and train neural network models in just a few lines of code.

5) Numpy: NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, fourier transform, and matrices. In this context it is mainly used for making information in images into data arrays.

6) Sklearn: Scikit-learn (Sklearn) is a useful and robust library for machine learning in Python. It provides a number of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistent interface in Python.

To reshape the array into required form and use it according to the needs of other libraries. MODEL TRAINING: Following parameters are selected in the project to train the model:

- 1) Data size: 22540
- 2) Number of neighbours: 5
- 3) Number of estimators: 200
- 4) Dimensionality reduction technique used: Principal component analysis
- 5) Classifier used: Voting classifier (KNN, Random Forest)
- 6) Testing data size: 20% of total dataset size
- 7) Model compilation based on accuracy, False positive rate, detection accuracy
- 8) Performance visualization by using confusion matrix
- 9) The time required for single sample is 0.0000384 secs

Confusion matrix:

