Senior Project: Character Recognition using Machine learning.

William Jessup University.

Dept. of Computer science

Piyush Bhadauriya

**Description:** -

Purpose of this project is to use machine learning algorithms for Handwritten character Recognition task. **Machine learning** algorithm is a computer program which learn from experience E with respect to some class of tasks T and performance measure P and its performance at tasks in T, as measured by P, improves with experience E. Character Recognition task in the context of this project refer to Identifying characters from Handwritten Character images. In Machine Learning we use data to train a Model for a Task. **Models** are like functions which map input x with output y. **Training** involves optimizing the model parameters based on the data so that it can be extended for other generalized input. Trained model then, can provide accurate output for new data. There are different types of training, this project uses Supervised learning to train the Machine learning model. **Supervised learning** means learning a function that maps from x to y, using labeled training examples (x,y). Supervised learning algorithms include linear regression, logistic regression, and neural networks.

OCR (**Optical Character Recognition**) technology to read a document and extract characters and words from it has been around for some time. But while OCR software are good at identifying printed character from the Documents, they perform poorly for the handwritten character. Main problem in identifying handwritten character is that every person has different writing style. It becomes really complex to use regular algorithms to account for different styles. Machine Learning algorithms are good solution for this kind of problems as they are designed to work in such complex scenarios. For character recognition applications, we need to select appropriate machine learning model as some models perform better then others for Specific tasks. We also need diverse and huge datasets to be able to train a good Model. Machine learning model uses data to learn so training data need to be carefully selected. If training data is missing for a specific group or type It can negatively impact the accuracy. for example, if our model never trained with data for some specific styles/font of handwritten characters then our model will perform poorly when used to predict characters written in that specific style.

Task to recognize a character from image can be modeled as a multiclass classification task. **Multiclass Classification** task involves calculating output y for different inputs x and using the model and classify them to different Categories based on the output. So, we need 10 separate categories (1-10) to identify numbers and 26 separate categories(A-Z) to identify English alphabets.

