Problem Statement and Goals
OCRacle

Phillip Tran

Table 1: Revision History

Date	Developer(s)	Change
2025-17-01	Phillip Tran	Document created

1 Problem Statement

1.1 Problem

Researchers analyzing physical print documents such as newspapers, books, and letters often need a means of digitizing the text in these documents. This enables them to search and analyze the text data more efficiently. Especially in the case of historical documents, digitizing the text can help preserve the information contained in these documents.

Optical Character Recognition (OCR) is a technology that allows for the extraction of text information from scanned documents, images, and other optical formats where text may be present. This digitalization process enables researchers to use computer programs to find trends and patterns in the digitized text.

1.2 Inputs and Outputs

Input: A black and white image containing a single Latin alphabet character to be recognized.

Output: The program's prediction of the Latin alphabet character in the image and its confidence level.

1.3 Stakeholders

A researcher interested in digitizing text from an image would be the most likely stakeholder for this tool.

1.4 Environment

The program will be compatible with Windows, MacOS, and Linux operating systems. Any modern computers capable of running the operating systems mentioned above should be able to run the program.

2 Goals

The program should be able to recognize Latin alphabet characters with an overall accuracy of at least 80%.

3 Stretch Goals

The program should be able to recognize Latin alphabet characters with an overall accuracy of at least 90%.

The program should be able to recognize Latin number characters with an overall accuracy of at least 80%.

4 Challenge Level and Extras

I expect this project to have a general challenge level. Although this project has been done before, I expect that it will be challenging to achieve a high level of accuracy in recognizing characters.

Since this is not considered a "research" project, I will be including a user manual as an extra. This will help users understand how to use the program and what to expect from it.