**Optical character recognition of handwritten Hebrew documents**

SE-D-10

Tomer Kakou; tomerkakou@gmail.com

Tal Bo-Ahron; talbo4@gmail.com

Advisor: Dr. Natalia Vanetik

SCE - Shamoon College of Engineering, Be’er-Sheva

As digital transformation accelerates, processing handwritten text images is increasingly crucial for searching, storage, and editing. While OCR for printed text has advanced, handwritten Hebrew remains challenging. We are developing a method integrating a curated dataset with the Hebrew HDD dataset for better classification. Our approach enhances image resolution and extracts character images using OpenCV. Each character is classified into 27 Hebrew letter classes, then reassembled into words. We use deep learning models like ViT and ResNet-50 for recognition, with LLMs providing contextual corrections. Evaluation metrics include character error rate (CER), word error rate (WER), and normalized Levenshtein distance (NLD). Our goal is an end-to-end Hebrew handwriting OCR solution with high accuracy, even for degraded text.

Keywords: Hebrew, image classification, low-resource languages, OCR.