

The title of my proposed research project is "Semi automated algorithm for the transcription of handwritten manuscripts". This project's purpose is to assist the transcription of medieval hand-written documents. Optical Characters Recognition (OCR) techniques are now widely spread and reliable for most modern fonts, but applying them to hand written text is still challenging. Historical and archeological research has always been data driven, and nowadays, data mining techniques offer new possibilities to explore those data. Ancient texts are a major source of data, but they first need to be deciphered, a long and error prone process, which can be addressed by algorithms to reduce the amount of time needed as much as the risk of errors.

My project will offer a tool to assist the deciphering of handwritten documents. Using classification algorithms, it is possible to assess the letters and determine, with various levels of accuracy, the handwritten characters. Under user guidance, the tool will expand its database and learn to assess the letters with more precision. The prediction's results will be enhanced by using statistical tools and by comparing the deciphered word with a lexical database, to determine how the assessed word can match a recorded word.