

Computational Paleography

Algorithms can be used for paleographic analysis in two very different ways, one human-oriented and the other machine-oriented, both of which will be explained and exemplified in this talk.

In a relatively straightforward fashion, algorithms can be designed to learn from experts which features of handwriting are of significance and how they are to be used for classifying handwritten texts and matching hands. In addition to the rules used for the determinations, this requires a large quantity of text tagged by humans for those handcrafted features. Once trained, the algorithm can analyze massive quantities of letters or letter combinations for the presence of these features, and determine provenance on that basis. Algorithms can also be used to measure the relative contribution of different features and optimize the decision process.

A more modern, machine-learning approach is to train neural networks on large amounts of manually classified images of texts, leaving it to the algorithm to discern relevant features, in much the same way that algorithms perform facial recognition. The salient features for the algorithm may have little in common with the paleographic features that underlie a paleographer's expertise. One potential shortcoming of this method is that it may be difficult for the algorithm to explain the reasons for its decisions. On the other hand, it may suggest novel features that could also serve expert paleographers.