A Bayesian Approach to Analysis of Handwritten Evidence

When document examiners conduct an authorship analysis of handwritten evidence they don't compare full pages of writing to full pages of writing. Instead, they often consider smaller, more manageable pieces such as letters. Likewise, our approach first breaks words in a document into smaller connected pieces of ink called "pseudo-letters." The frequency at which pseudo-letters appear for a writer and measurements taken on those pseudo-letters serve as data for Bayesian hierarchical modeling and Bayesian Additive Regression Trees (BART). Letting writers with differing styles (cursive, print, etc.) cluster into smaller, more homogeneous groups allows us to assign different priors to similar groups of writers within the hierarchical structures. Using these methods we are able to easily identify authors of questioned writing in a closed set of writers using only a handful of pseudo-letters and measurements. We discuss results of an analysis on an open set of writers, which is a more realistic scenario when dealing with problems in forensic evidence. Our procedure is context (and language) independent while remaining interpretable, since pseudo-letters often correspond to roman letters.