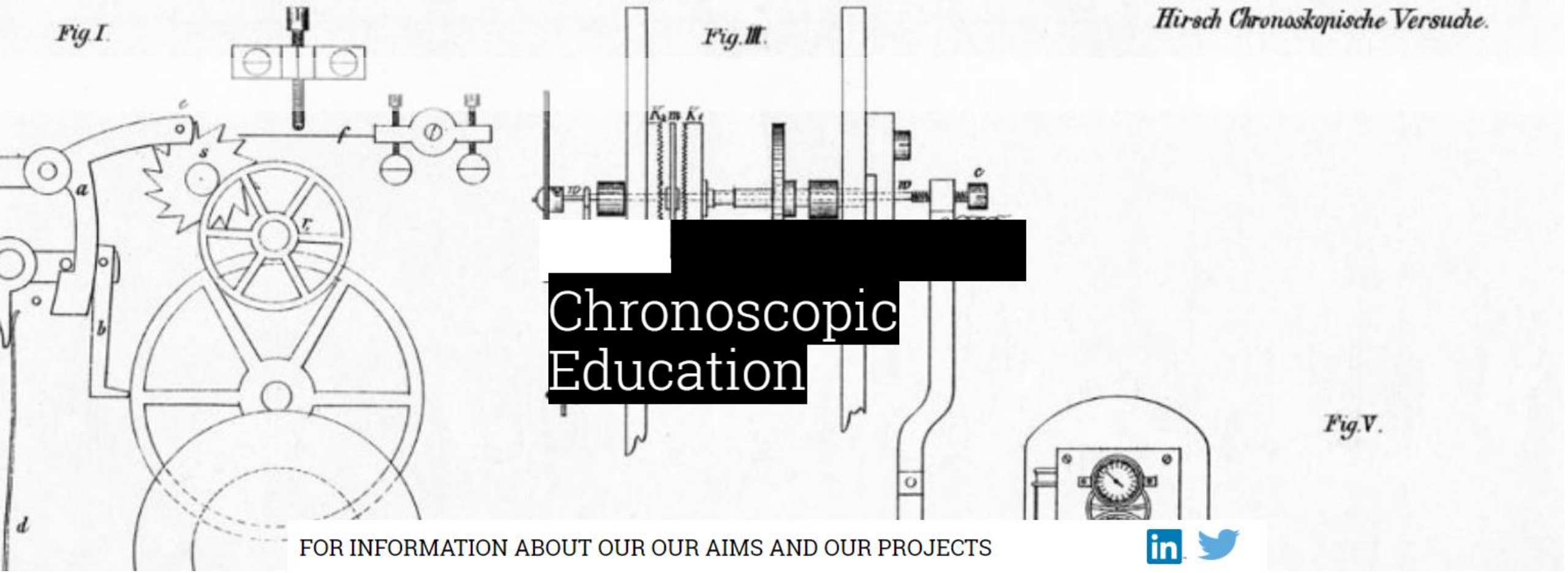


Chronoscopic Education

FOR INFORMATION ABOUT OUR OUR AIMS AND OUR PROJECTS

Digital skills needs amongst historians

Discussion with Professor Jane Winters

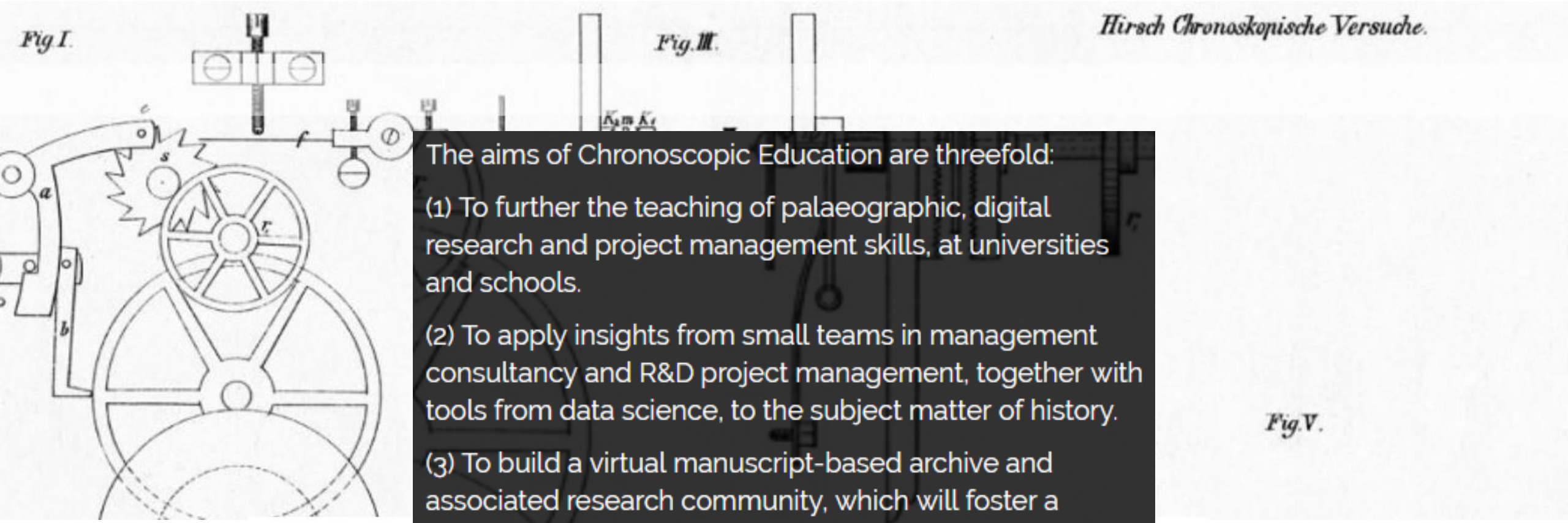
Colin Greenstreet

Wednesday, July 25th 2018

Discussion topics

- Current state of digital training for historians in higher education
- Anticipated skill needs in next five years
  - To meet changing needs of non-academic employment market
  - To ensure requisite skills for funded PhDs and post-docs
  - To ensure requisite skills for junior academic positions in history
- Chronoscopic Education vision for digital training
  - Integrating concepts, data, tools and coding
  - Integrating academia with GLAM & public history
  - Using MarineLives, Signs of Literacy, SOLM-2018, Glossary & Gazetter content
  - Using facilitated small teams
  - Data management, database design, semantic annotation, wikidata, linked open data, image extraction
  - Project management, GitHub

Our social aims



The aims of Chronoscopic Education are threefold:

(1) To further the teaching of palaeographic, digital research and project management skills, at universities and schools.

(2) To apply insights from small teams in management consultancy and R&D project management, together with tools from data science, to the subject matter of history.

(3) To build a virtual manuscript-based archive and associated research community, which will foster a culture of collaborative scholarship.

http://www.chronoscopic.org

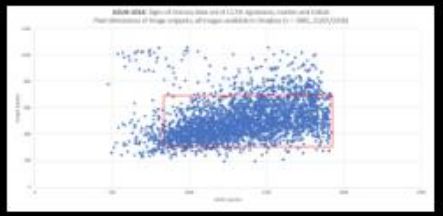
Project portfolio

MarineLives

Maphackathon

Signs of Literacy

SOLM-2018



EM Textiles, Garments & Dyestuffs Glossary

EM Maritime & Mercantile Gazetteer

http://www.chronoscopic.org

Signs of Literacy Kaggle Research Competition, Nov 2018 – Jan 2019

Google owned Kaggle has selected us as one of a small number of pro bono competitions they support each year on the merits of our proposal, and the potential impact on the research field and community of the competition.

Kaggle will cover the running costs of the competition. We will provide the prize pool, and are now seeking to raise US $30,000 from potential sponsors and partners.

The Proof of Concept will contain two parts:

(1) Algorithmic identification of markes, initials and signatures.

(2) Algorithmic discrimination between degrees of "sophistication" within the three categories of "marke"; "initial(s)", and "signature".

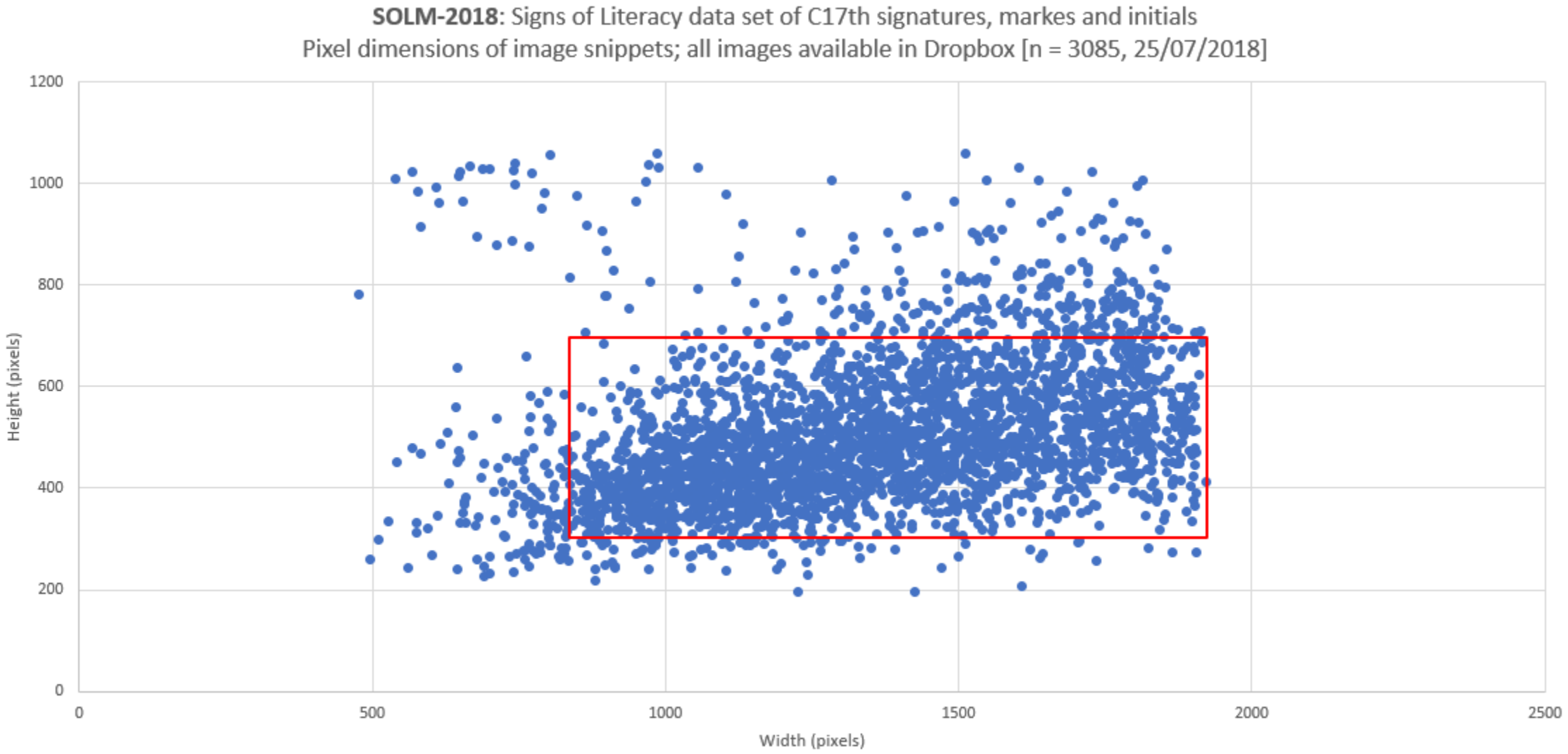
Having proven the concept, we will seek out an image or vision oriented computational laboratory with which to develop a grant funded collaboration to take the work further in 2019 and beyond.

https://www.linkedin.com/pulse/proposed-signs-literacy-kaggle-research-competition-2018-greenstreet/

Technical vision & role of the Kaggle competition

- Automatic identification of manuscript pages containing signoffs
- Markup of manuscript pages to isolate signoffs (markes, initials, signatures)
  - Hand markup of manuscript pages, but ideally automated markup
  - Signoffs can be single or multiple, for single or multiple depositions
  - Deponent signoffs; interpreter signoffs
- Automatic differentiation between classes of markes, initials and signatures
- Automatic differentiation within each class as to sophistication of execution & other parameters as a surrogate for literacy
- Automatic identification of manuscript pages containing deponent metadata (name; age; occupation; place of residence; date of deposition)
- Markup of manuscript pages to isolate deponent metadata
- Hand writing text recognition of deponent metadata and associate metadata with correct signoff

Signs of Literacy Kaggle Research Competition, Nov 2018 – Jan 2019

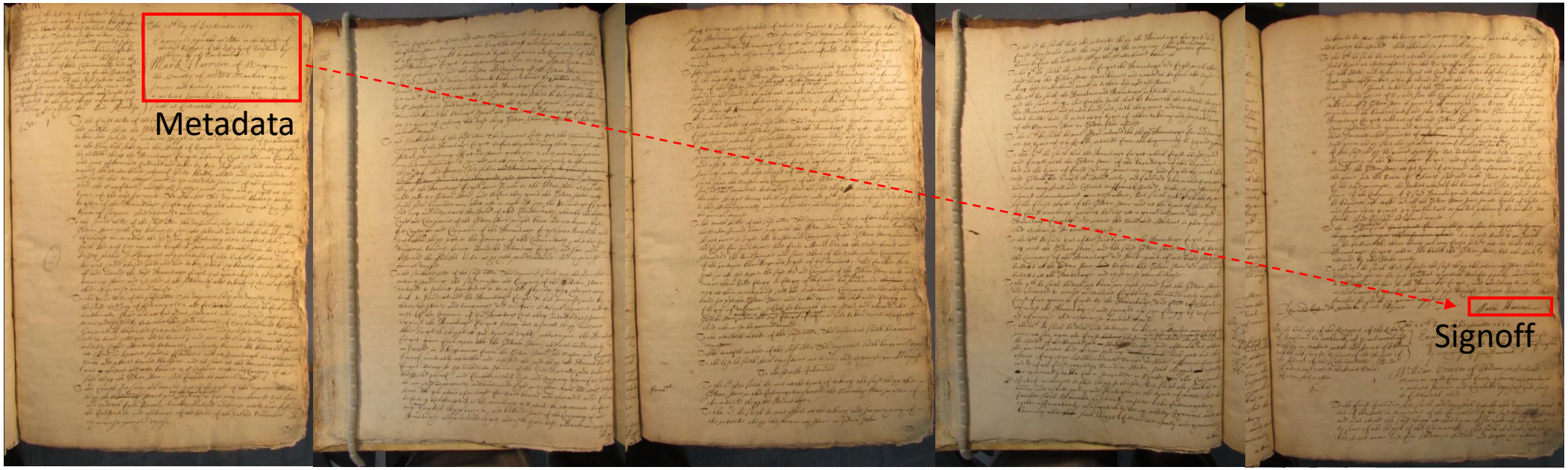


SOLM-2018 database

SUPPLEMENTARY MATERIAL

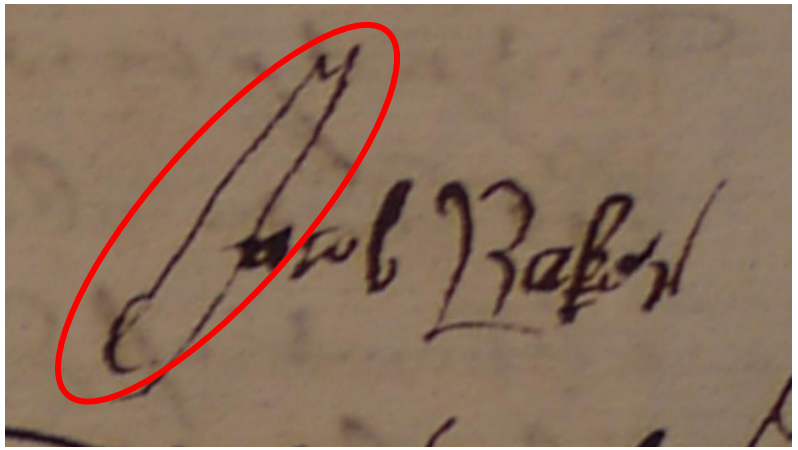
Legal deposition

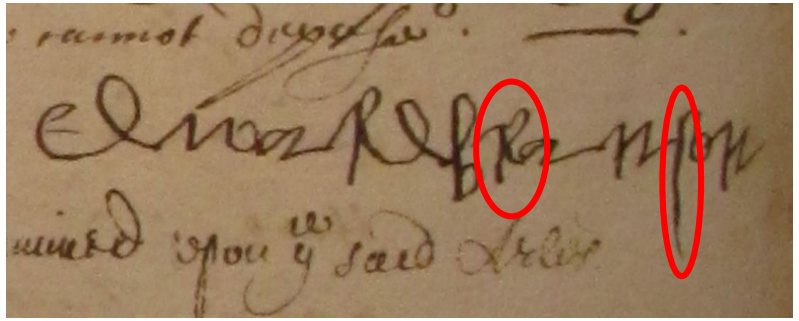
Deposition of Mark Harrison; mariner and master; resident in Wapping, Middlesex; age 27;
Dated September 21st 1659 (TNA, HCA 13/68, ff. 1r-3r)

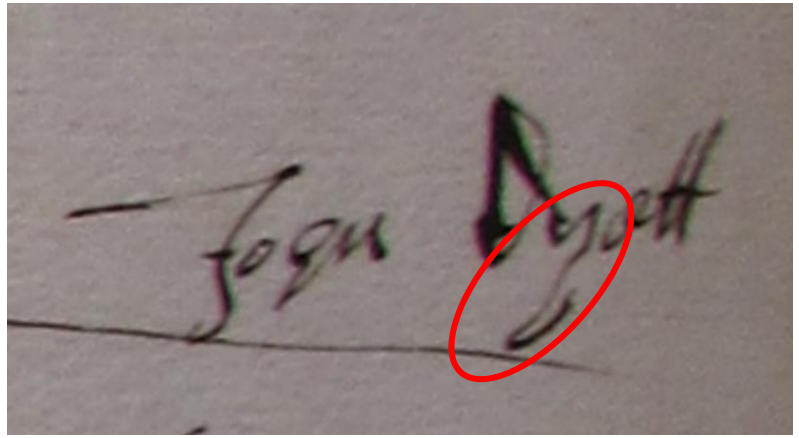


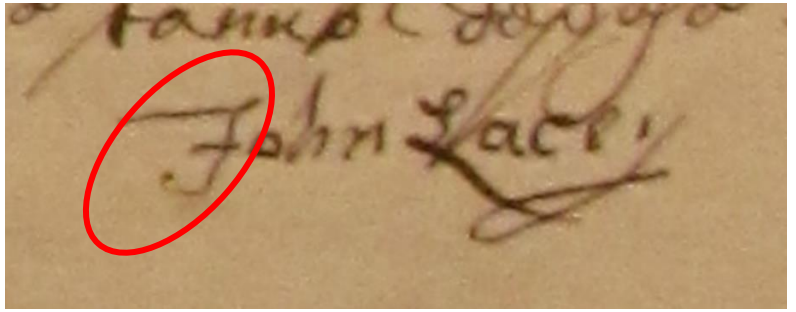
http://www.marinelives.org/wiki/HCA_13/68_f.1r_Annotate; http://www.marinelives.org/wiki/HCA_13/68_f.1v_Annotate; http://www.marinelives.org/wiki/HCA_13/68_f.2r_Annotate ;
http://www.marinelives.org/wiki/HCA_13/68_f.2v_Annotate; http://www.marinelives.org/wiki/HCA_13/68_f.3r_Annotate

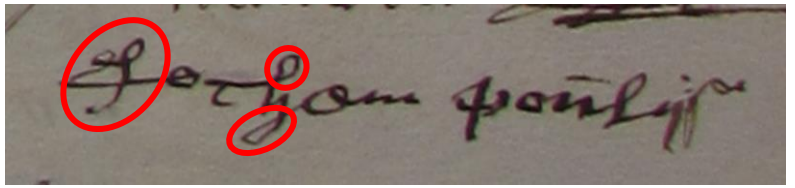
Physical characteristics of poorly executed signatures for machine detection – Part 1

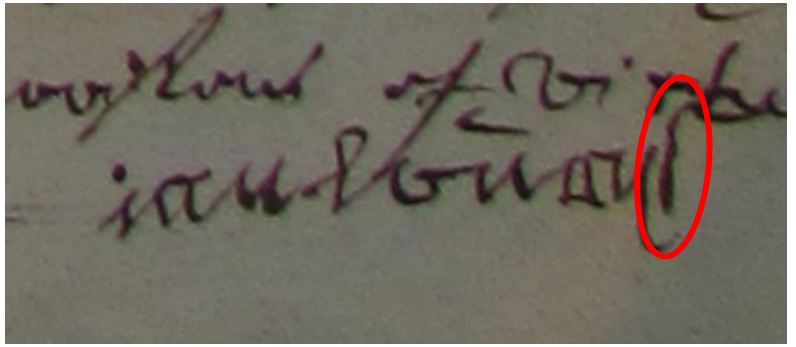








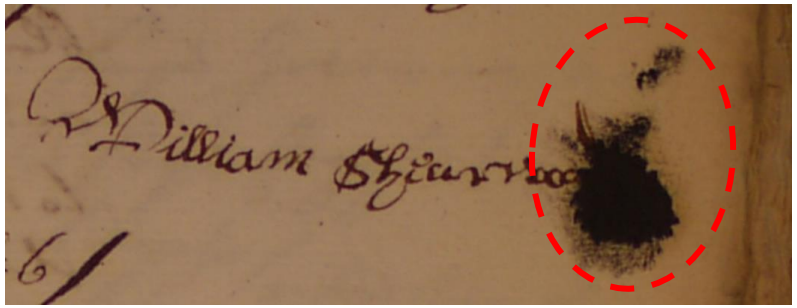


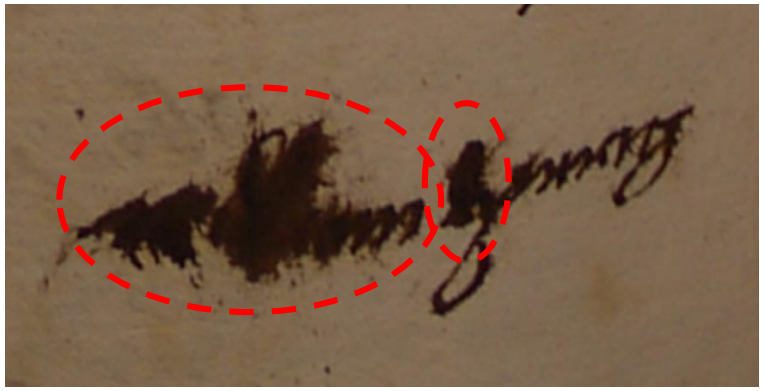


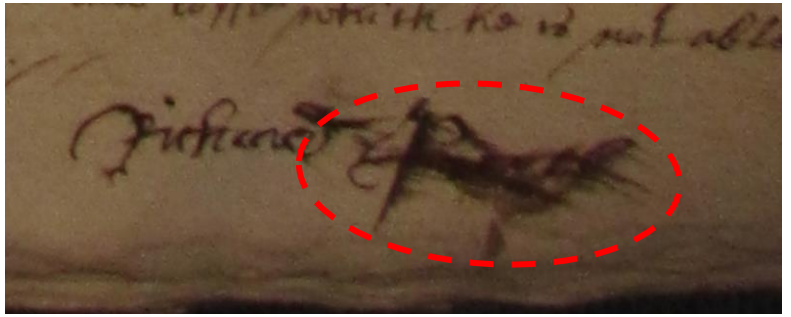
Source: Clockwise from top LH side:
KaggleTestSnippet_HCA_1371_f.263v.PNG,
KaggleTestSnippet_HCA_1368_f.483v.PNG,
KaggleTestSnippet_HCA_1368_f.51v.PNG,
KaggleTestSnippet_HCA_1370_f.20v.PNG,
KaggleTestSnippet_HCA_1370_f.22r_Two,
KaggleTestSnippet_HCA_1370_f.168v.PNG

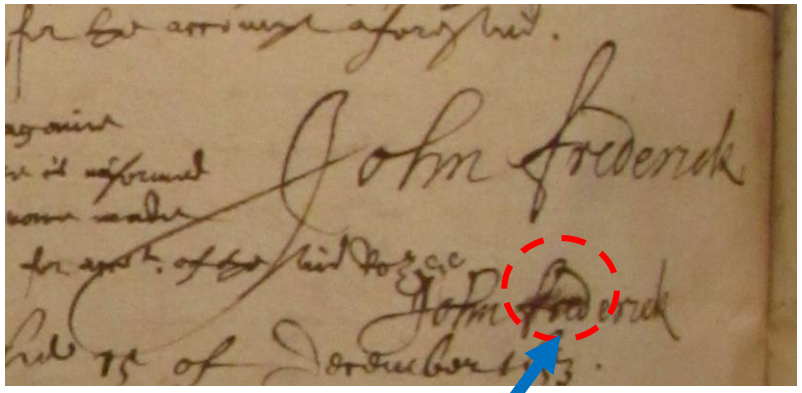
Shaky straight lines and/or loops

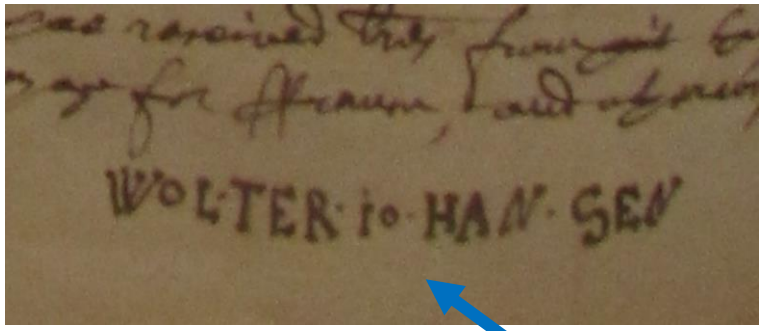
Physical characteristics of poorly executed signatures for machine detection – Part 2

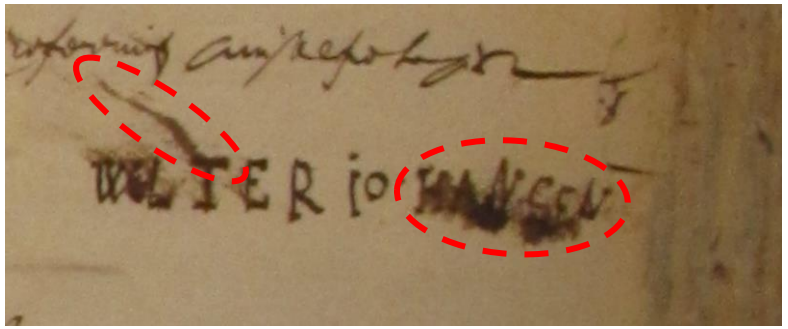












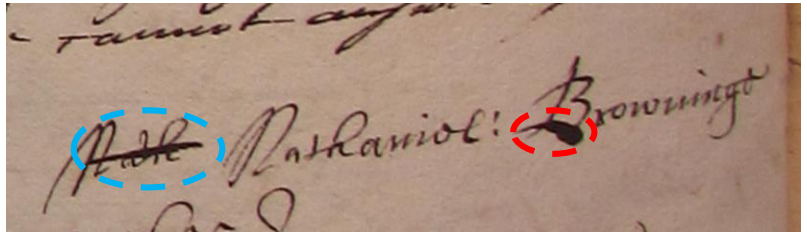
Even London alderman & merchant, John Frederick, could smudge his signature, when signing an addendum to his deposition three weeks after the first deposition

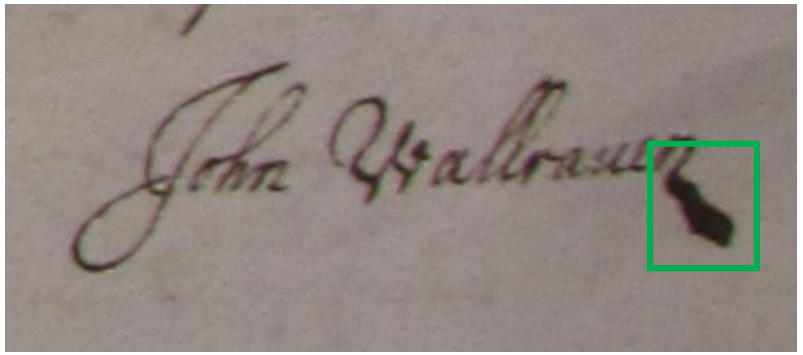
Wolter Johansen signed a second deposition smudge free, whilst signing both times with capitals

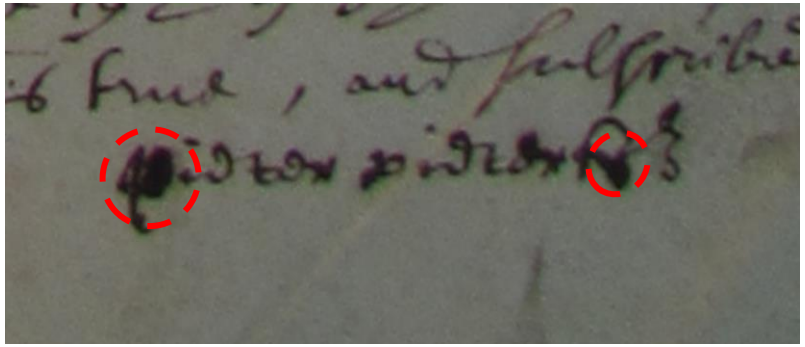
Ink blots or smudges

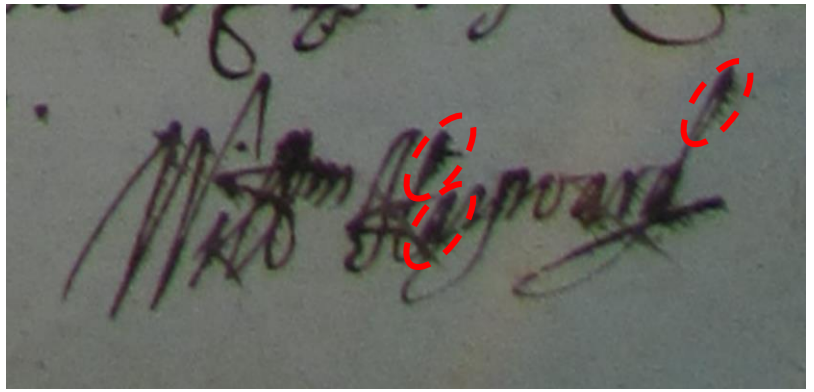
Source: Clockwise from top LH side: KaggleTestSnippet_HCA_1371_f.503r.PNG, KaggleTestSnippet_HCA_1373_f.498v.PNG, KaggleTestSnippet_HCA_1368_f.59r.PNG, KaggleTestSnippet_HCA_1368_f.231r.PNG, KaggleTestSnippet_HCA_1368_f.239v.PNG, KaggleTestSnippet_HCA_1368_f.241v.PNG

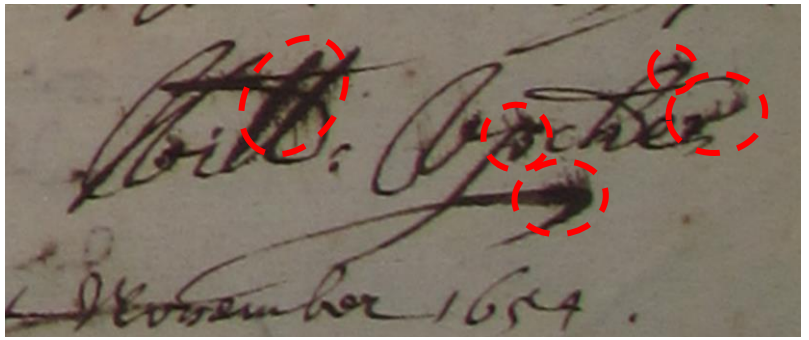
Can machine detection distinguish blots, smudges, stylistic features, & deletions? (1)

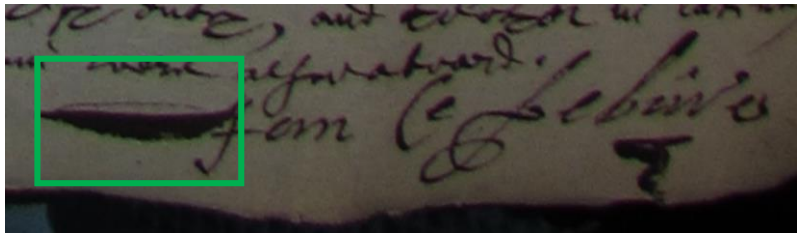




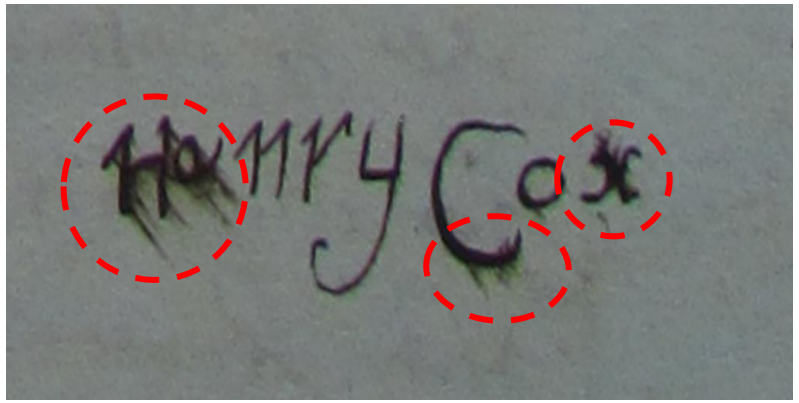


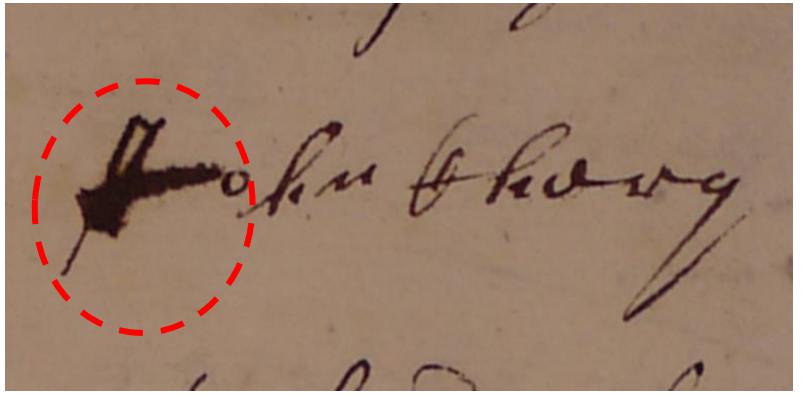






Source: Clockwise from top LH side:
KaggleTestSnippet_HCA_1370_f.387v.PNG,
KaggleTestSnippet_HCA_1370_f.13r.PNG,
KaggleTestSnippet_HCA_1370_f.167r.PNG,
KaggleTestSnippet_HCA_1371_f.456r.PNG,
KaggleTestSnippet_HCA_1370_f.15r.PNG,
KaggleTestSnippet_HCA_1370_f.19r.PNG,
KaggleTestSnippet_HCA_1370_f.41v.PNG,
KaggleTestSnippet_HCA_1370_f.17v.PNG,



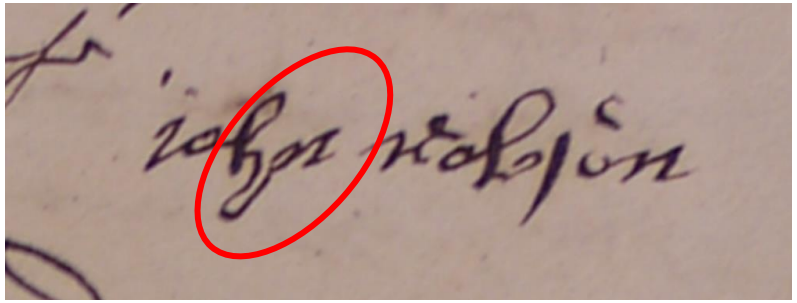


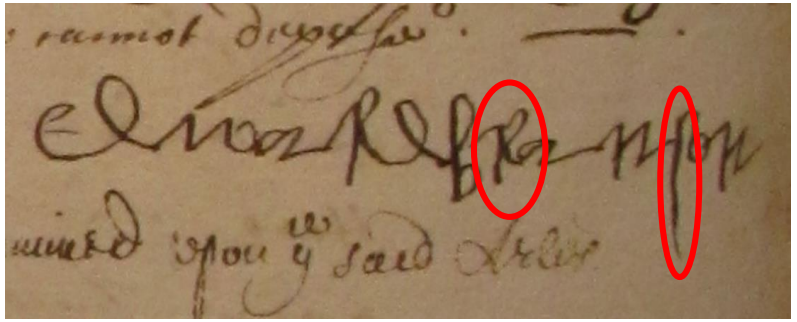
Ink blots or smudges

Stylistic feature or smudge?

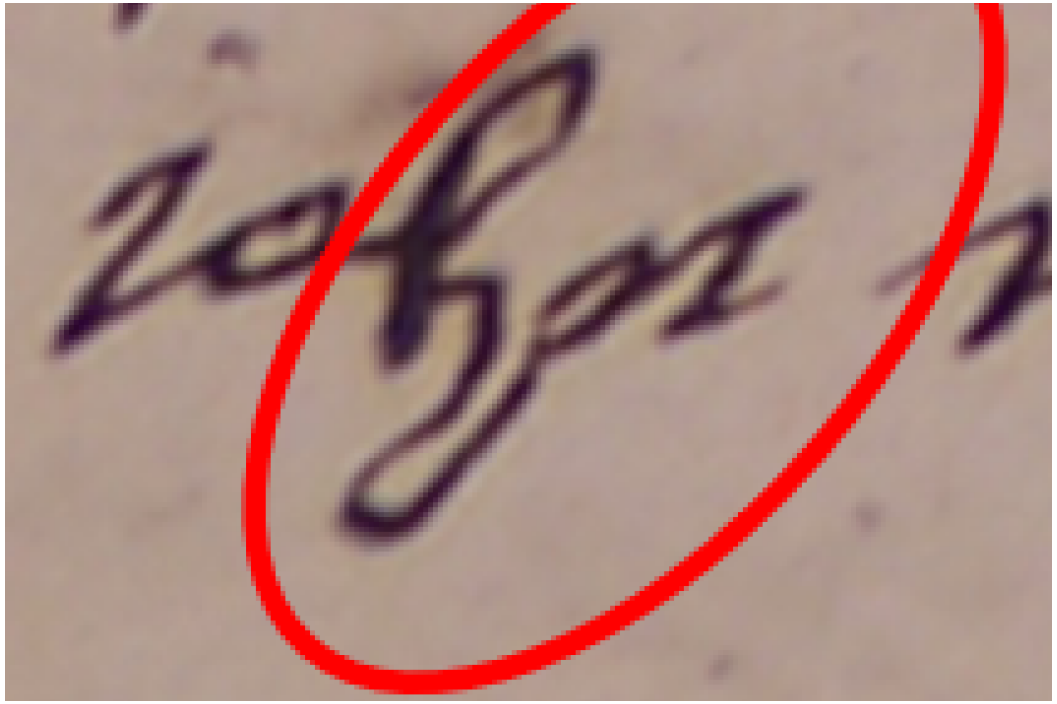
Deletion

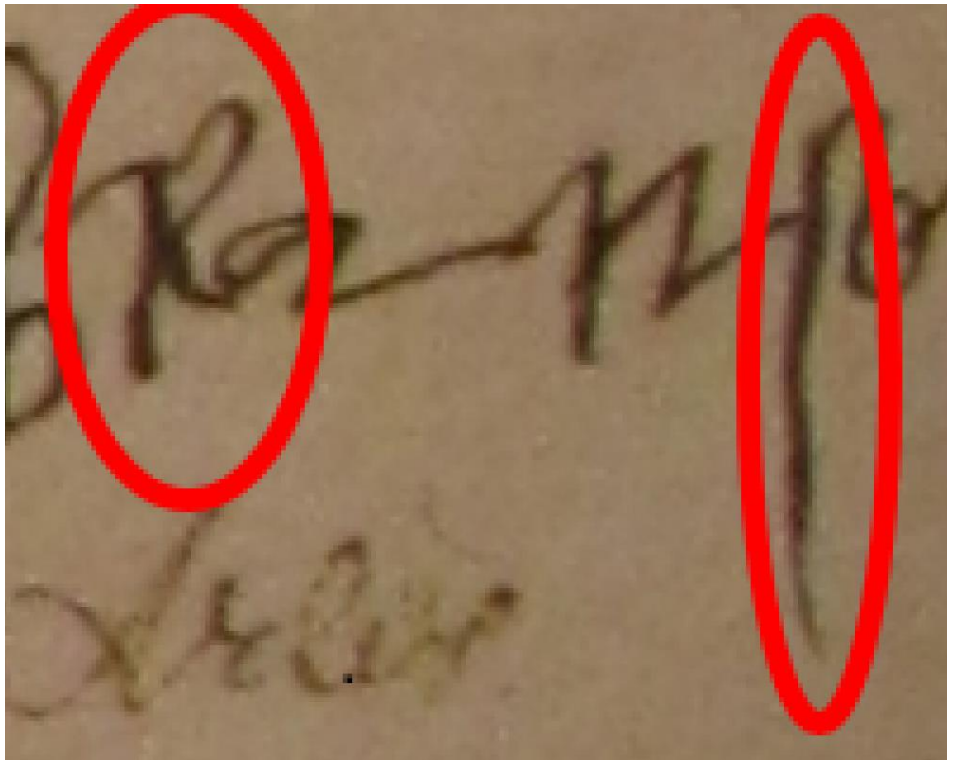
We are looking for algorithms to detect “shake” in straight and curved lines





HYPOTHESIS: Shaky lines are a sign of poor signature execution (and by inference, poor handwriting execution) suggesting lower level of literacy than smooth executed lines





Shaky straight lines and/or loops

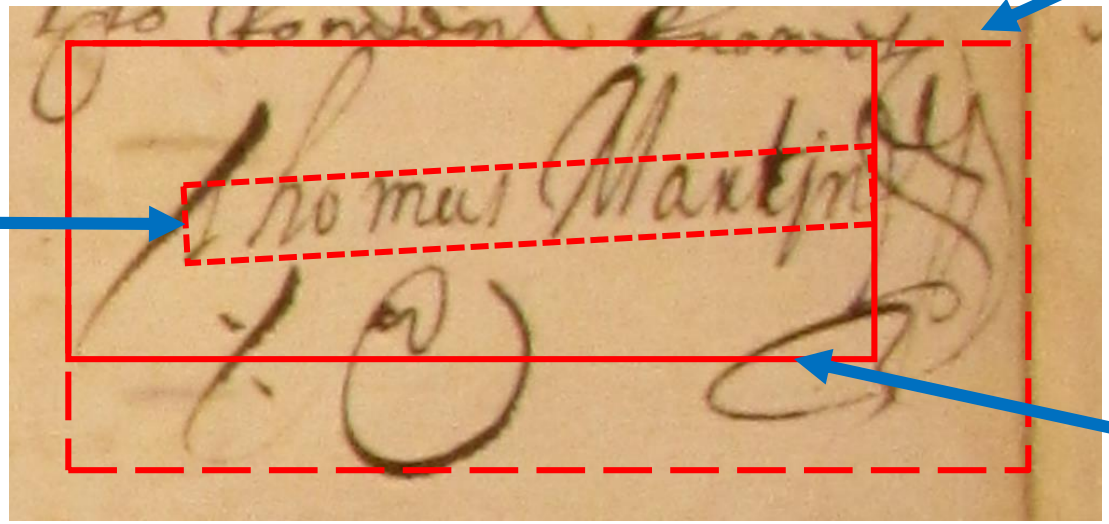
Clockwise, from top LH: KaggleTestSnippet_HCA_1371_f.435v.PNG, KaggleTestSnippet_HCA_1368_f.483v.PNG, KaggleTestSnippet_HCA_1368_f.483v_PIXELS.PNG, KaggleTestSnippet_HCA_1371_f.435v.PNG_PIXELS.PNG

Putting boundary boxes on C17th signatures

Boundary boxes marking the visual geometry of a signature

Outside boundary box, including flourish

Inside boundary box, excluding uppers and downers

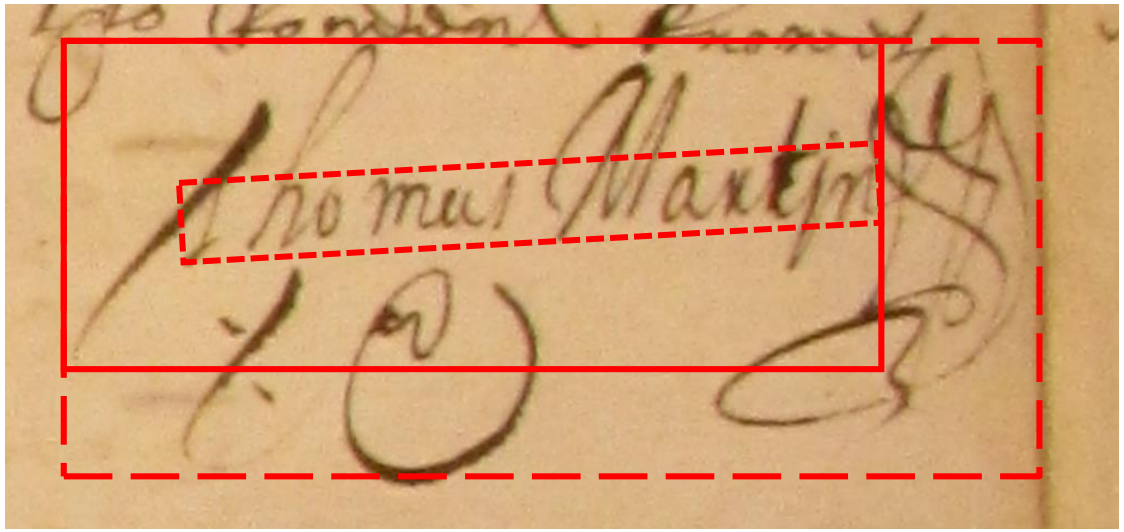


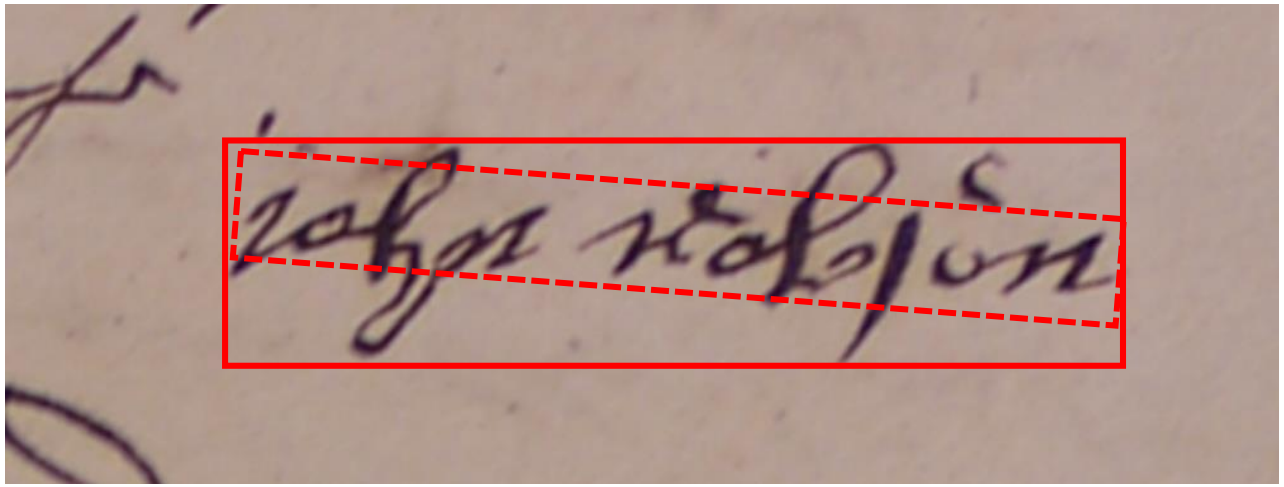
Middle boundary box, including all letters, but excluding flourish

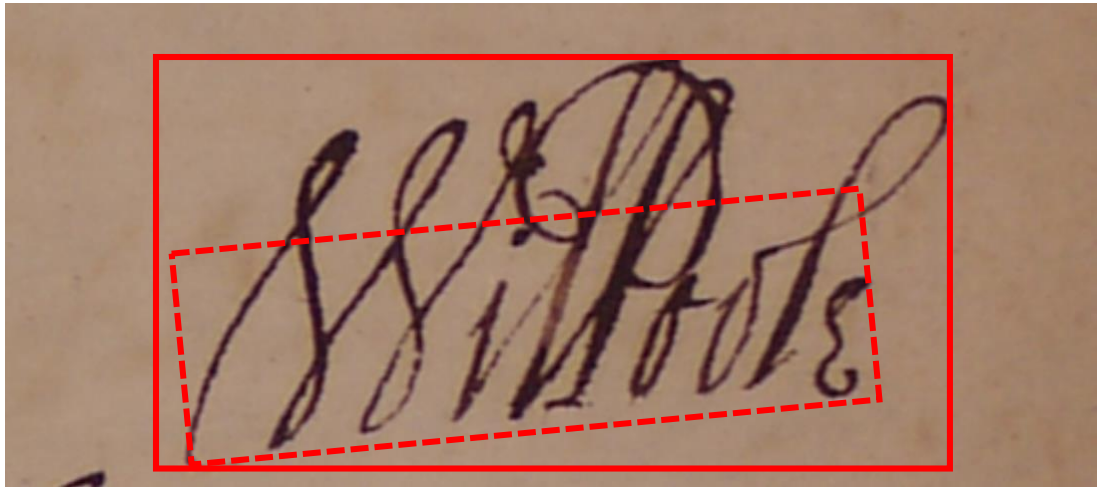
Statistics

Inside boundary box: 9.0 x 1.1
Middle boundary box: 9.75 x 4.25
Outside boundary box: 12.75 x 5.75
Rotation from horizontal: ca. 340 degrees

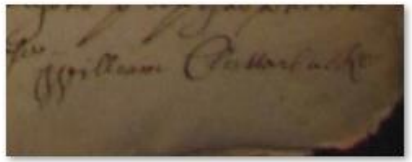
Different visual geometries of signatures

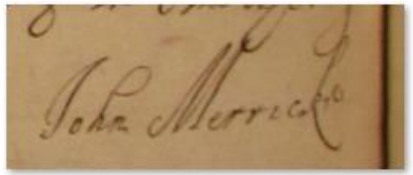


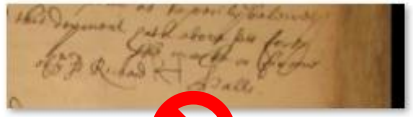


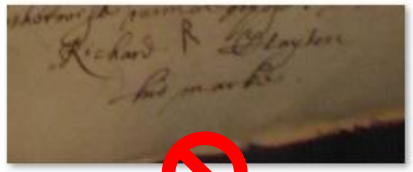


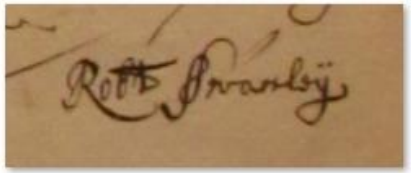
Challenge One: Identify the 11 image snippets on this page which contain signatures and highlight the 2 image snippets which contain signatures belonging to the same person

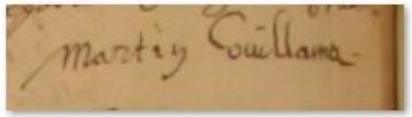

KaggleTestSnippet_HCA_1368_f.140r.PNG


KaggleTestSnippet_HCA_1368_f.142r.PNG


KaggleTestSnippet_HCA_1368_f.143r.PNG


KaggleTestSnippet_HCA_1368_f.145r.PNG

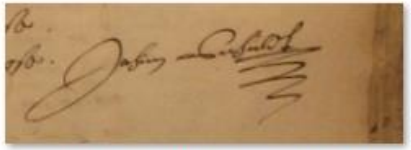

KaggleTestSnippet_HCA_1368_f.148r.PNG

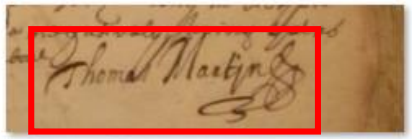

KaggleTestSnippet_HCA_1368_f.150v.PNG

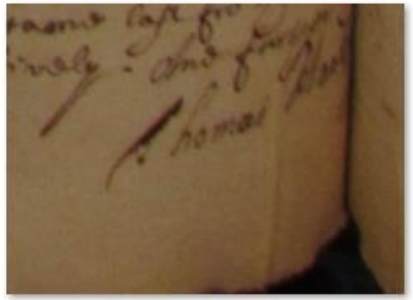
KaggleTestSnippet_HCA_1368_f.151v.PNG

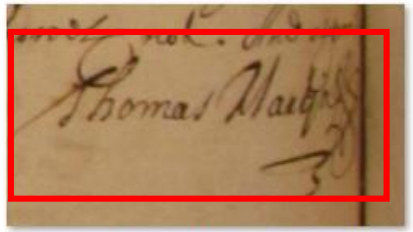
KaggleTestSnippet_HCA_1368_f.153v.PNG

KaggleTestSnippet_HCA_1368_f.155r.PNG


KaggleTestSnippet_HCA_1368_f.156r.PNG


KaggleTestSnippet_HCA_1368_f.158r.PNG


KaggleTestSnippet_HCA_1368_f.159v.PNG

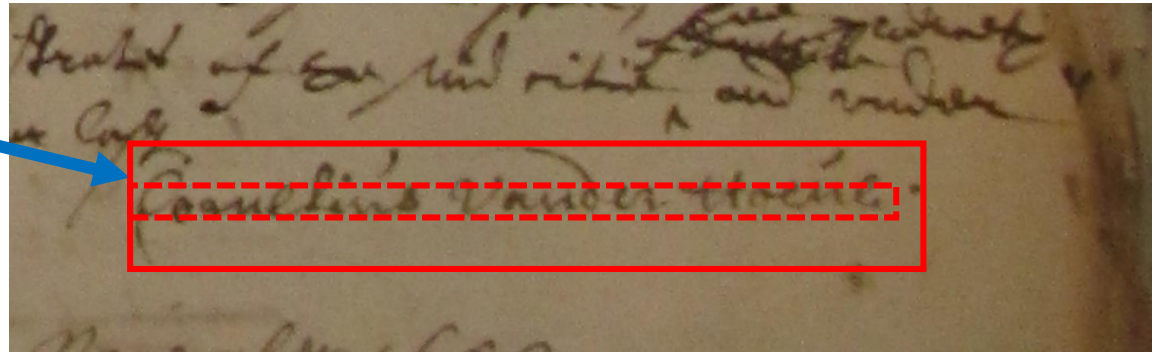

KaggleTestSnippet_HCA_1368_f.161v.PNG

Source: KaggleTestSnippets data set

Challenge Two: Detect a London based merchant, who has Dutch origin, from physical characteristics of signature, rather than spelling of name

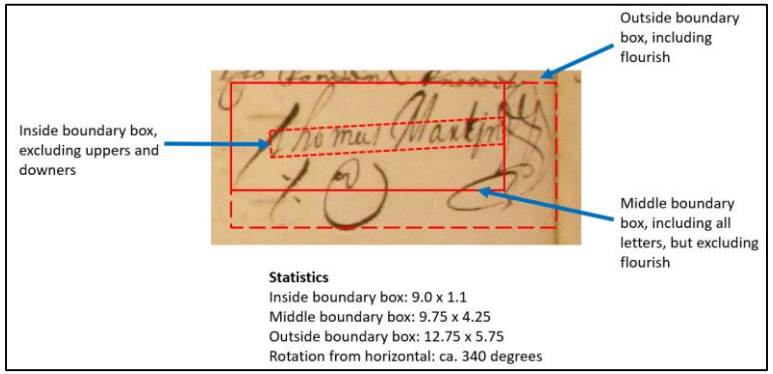
Semantic giveaways

- 3 names, not 2
- Specific names [“Cornelius”; “Vander”; “Hoene”]



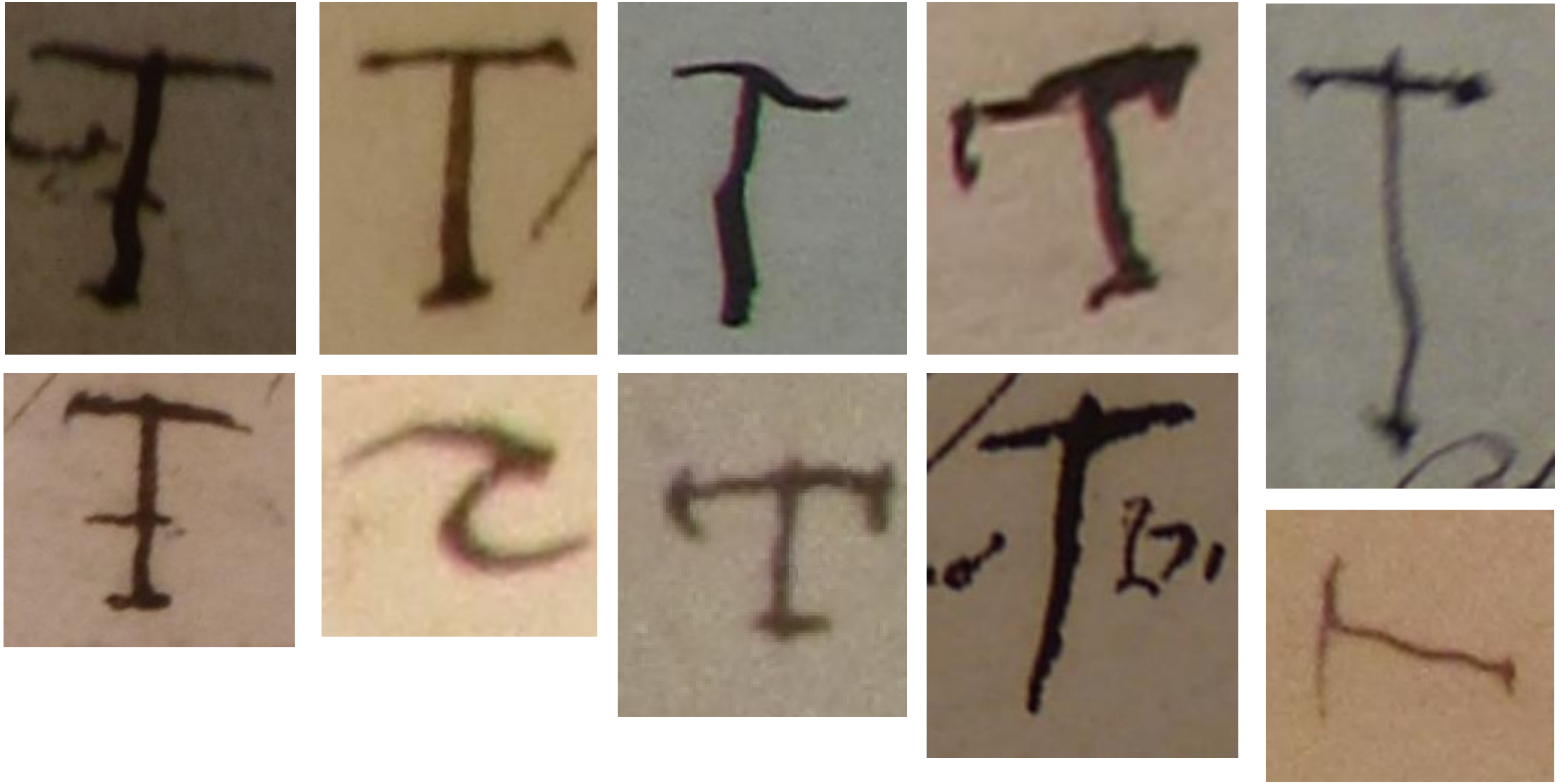
Physical giveaways

- Long, narrow signature
- No flourish
- Limited capitalisation
- Middle bounding box close to inner bounding box, with no outside bounding box



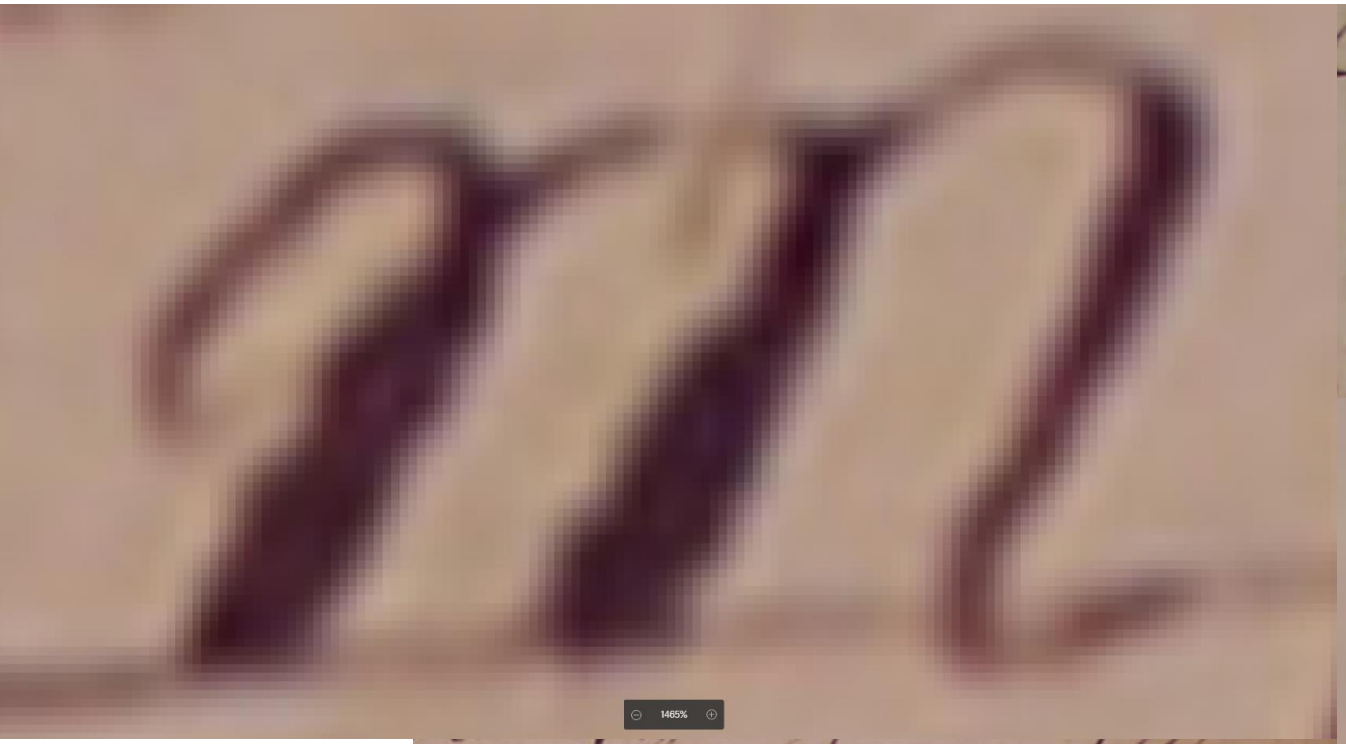
Source: KaggleTestSnippet_HCA_1368_f.233r_Two.PNG

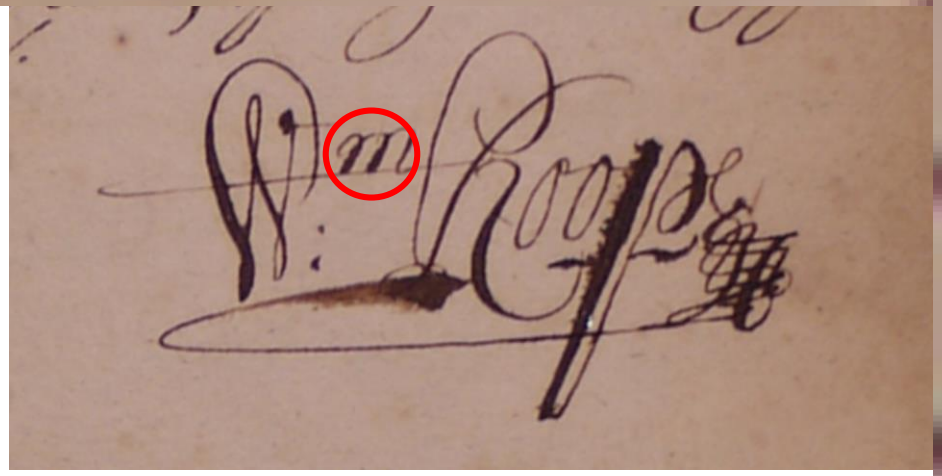
Research question - methodology: Can a machine distinguish between the authors of a specific initial?

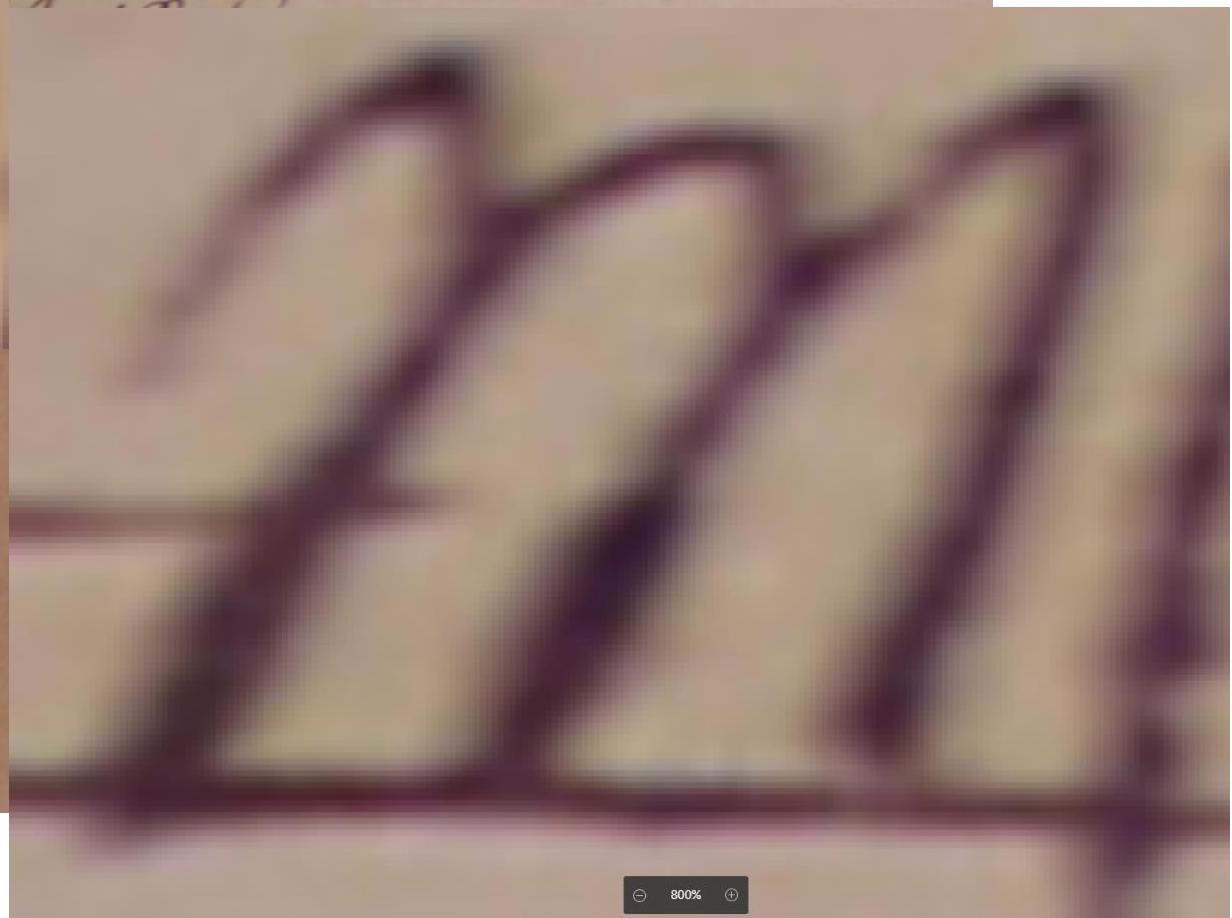


Source: KaggleTestSnippets

Research question - methodology: Measuring pixel dimensions







Source: KaggleTestSnippet_HCA_1373_f.199r.PNG, KaggleTestSnippet_HCA_1373_f.199v_One.PNG