CA Journal of Cultural Analytics

Detecting Footnotes in 32 million pages of ECCO

Sherif Abuelwafa, Sara Zhalepour, Ehsan Arabnejad, Mohamed Mhiri, Emilienne Greenfield, James P. Ascher, Sofia Bach, Victoria Svaikovsky, Alayne Moody, Andrew Piper, Chad Wellmon, and Mohamed Cheriet

12.03.18

Peer-Reviewed By: Anupam Basu

Clusters: Data, Image

Article DOI: 10.22148/16.029

Dataverse DOI: 10.7910/DVN/FMZYFP

Journal ISSN: 2371-4549

Cite: Sherif Abuelwafa, Sara Zhalepour, Ehsan Arabnejad, Mohamed Mhiri, Emilienne Greenfield, James P. Ascher, Sofia Bach, Victoria Svaikovsky, Alayne Moody, Andrew Piper, Chad Wellmon, and Mohamed Cheriet, "Detecting Footnotes in 32 million pages of ECCO," Journal of Cultural Analytics. *December 3, 2018.*

In "An Answer to the Question: What is Enlightenment?", the eighteenth-century German philosopher Immanuel Kant responded to a big question buried in a little footnote. But you wouldn't know it, because contemporary editions of Kant's famous essay no longer reproduce the parenthetical directive that Kant's original essay printed right under the essay's title in the December issue of the Berlinische Monatsschrift in 1784: "S. Decemb. 1783. S. 516." (See December 1783, p. 516). And, in fact, page 516 in the December volume of the Berlinische Monatsschrift 1783 has a footnote: "What is Enlightenment? This question is nearly as important as: what is truth? And should certainly be answered before one starts to enlighten! But I have yet to find it answered anywhere."

Kant's attempt to define enlightenment, then, was a reply to a specific question. The footnote to which Kant's essay refers was published in an essay written by the Berlin pastor Johann Friedrich Zöllner, who had published several sermons in the Berlinische Monatsschrift. Zöllner's essay, "Is it wise to no longer sanction marriage through religion?", discussed whether it would be "enlightened" to no longer require clergy to officiate weddings.[1] Kant's essay was addressed to a larger philosophical question, but also a particular question posed on a particular page in a particular periodical by a particular contemporary. And yet we tend to read Kant's essay, and, thus, the Enlightenment, apart from these relationships and particular moments of printed address and response.

Footnotes like the one on the bottom of page 516 in Zöllner's essay are visible traces of these mediated relationships, markers of conversations, references, connections, and the sociability of knowledge. They are also visible markers of legitimacy and authority. They demonstrate familiarity, identity, and knowledge. As Anthony Grafton writes with respect to the eighteenth-century origins of the footnote within the nascent discipline of history: "The footnote is bound up in modern life with the ideology and the technical practices of a profession."[2] It is an element of the history of disciplinarity and scientific credibility.

The footnote is also one of several visual typographic practices that have shaped modern knowledge. Our effort to understand better the footnote's place within the eighteenth century is part of a larger, on-going project that seeks to understand a range of visual practices of scientific notation in the past: whether it be footnotes that communicate authority and the relationality of sources; tables that bring together disparate forms of information into geometric relations; diagrams that provide abstract representations of intellectual procedures or natural phenomena; or illustrations that provide mimetic representations of objects in the world. In each of these cases, authors, editors, and publishers used a graphic process to convey information and make truth claims, often in a way that sought to reduce complexity. And contemporary scholars continue to use these processes in order to communicate well and more efficiently with one another. Instead of reproducing the entirety of another text, we cite it. Instead of reproducing all of the underlying data of a process, we transform it into a table of relations. Diagrams abstract more detailed processes into more formal essence. Even illus-

[1] Pasanek, Brad and Chad Wellmon. "Enlightenment, Some Assembly Required." In *The Eighteenth Centuries: Global Networks of Enlightenment*, eds. David T. Gies and Cynthia Wall (University of Virginia Press, 2018).

[2] Anthony Grafton, *The Footnote: A Curious History* (Cambridge: Harvard UP, 1999).

trations have an indexical relationship to the larger real-world phenomena they are meant to represent.

As we explain in greater detail elsewhere, our larger project is about bringing together the intersecting strands of research from the fields of book history, the history of science, and document image analysis to better understand the analytical unit called "the page image" and its role in the history of scientific knowledge. Our aim us to take seriously the page image in a double sense: first, as an image of a page, that is, to see the digitized page first and foremost as an image rather than a flawed mediation of text; and second to see the page itself as an image, as a visual unit rather than a primarily textual one. What have been the ways that the graphic practices of pages have underpinned the epistemic claims of scientific knowledge?

In this essay, we recount our process of using machine learning and classification algorithms to detect footnotes within the Eighteenth-Century Collections Online dataset (ECCO). ECCO represents one of the most complete digitized collections of a national publishing context within a specific historical period, consisting of over 100,000 volumes and 32 million pages published in Britain between 1700 and 1800. It has become a staple of research in the history of ideas, not just in Britain but for scholars of the Enlightenment more generally. We see the enrichment of collections like ECCO as a primary research goal for furthering historical understanding.

We discuss here the samples of training data that were collected and manually annotated, the different types of page-features that were used in the detection process, and the estimated accuracy of our predictions. The net result is metadata on the presence of footnotes within approximately thirty-two million pages of historical documents, which we share along with metadata regarding the initial training data used so that others can work with the same data. As we detail in Table 1, overall we are able to recall pages with footnotes with 67.87% accuracy and of those we achieve a precision of 96.2%. This suggests that there are a considerable number of footnotes we may be missing but that when we do detect them we do so with a very high degree of confidence. In addition to these summary statistics, we also provide users with an estimated percentage of footnoted pages per document, a table of all page IDs that have predicted footnotes on them, and finally the estimated probability of a footnote being present for all pages in ECCO. We see this as a first step in fully annotating ECCO according to our four visual categories of footnotes, tables, diagrams, and illustrations.

We want to emphasize from the outset just how challenging this process has been. "At first glance, all footnotes look very much alike," writes Grafton, "[but] even a brief exercise in comparison reveals a staggering range of divergent practices." As we quickly learned, discerning what constitutes a footnote in the eighteenth-century is by no means a straightforward process. Training machines to identify such visual ambiguity is even more difficult. One of the issues that will need further reflection are the trade-offs between the gains of acquiring knowledge at larger scale and the challenges introduced by a fundamental uncertainty surrounding historical evidence.

Overall, we see this project contributing to a larger effort of enriching digitized collections of historical documents with more information about the documents (what is traditionally called "metadata" or data about data). We see this particular effort as contributing to knowledge about the visual qualities of page images, with a specific attention to what we are calling the history of "scientific notation." One of the major obstacles for historical understanding is the minimal amount of knowledge we have about individual documents within large document collections. We might think of this as a second-wave of digital history: the first—which is still on-going—involves the act of digitization itself. This effort is about making physical copies, which are geographically limited in their accessibility, more widely accessible to a broader reading public. The second wave, to which we see our work contributing, can be seen as the attempt to provide more knowledge about the composition of the collections to facilitate large-scale study of cultural history. ECCO metadata currently consists of features like publication date, author, title, publisher, and in some cases subject headings. There is much more that we can do to annotate collections. But to do so at large-scale requires developing algorithmic procedures for expertly labeling documents, which in turn necessitates greater collaboration between the humanities and the sciences.

However, it is important not to mistake these labels for what computer scientists call "ground truth." All knowledge is situated. This project represents the coordinated efforts of a team of 14 researchers split between the humanities and computer science, including both students and faculty, ranging in levels from BA to Masters to Phd to Professor. It entailed a lengthy collaboration to create mutual understanding and shared goals as well as a clear understanding of the cultural object of study (in this case historical practices of footnotes). The training data assembled thus represents the understanding and prior knowledge of the humaities cohort, while the detection algorithms represent the understanding and prior knowledge of the computer science cohort. Any machine learning process inevitably encodes, explicitly and implicitly, these biases into its outputs and are thus not value free. While this may seem less controversial with a more straightforward visual object like footnotes, it is important that we continue to foreground the human learning behind machine learning. We now proceed to describe the process we used to annotate thirty-two million pages of ECCO.

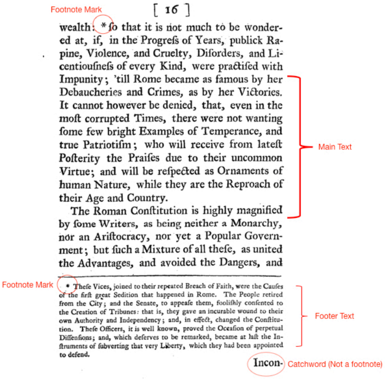
What is a footnote? (Training Data)

We began our research by defining a footnote and then identifying positive and negative examples within ECCO. For our purposes we defined a footnote as:

> **Footnotes need to be distinct, marked text at the bottom (foot) of the page that are referenced in the main part of the text.**

Each of these components is important: footnotes have a distinct location; they are marked (i.e. have a distinct marker); and refer directly to a location within the main body of the text through a matching mark (Fig. 1). Such a definition rules out side-notations (Fig. 2) or unmarked commentary that may be located at the bottom of the page (Fig. 3). Footnotes require some rule-based distinction of being "off-set". Despite these clarifications, we encountered numerous examples of pages that looked deceptively like footnotes (Fig. 4, Fig. 5). Because footnote marks are both highly varied in the eighteenth century and also highly indistinct as images (what is the difference between a poorly printed asterisk and an ink blot from the reproduction process?) (Fig. 6), the footnote mark, as we quickly learned, is only weakly significant in discriminating between footnotes and commentary. For this reason, we encountered a serious limitation in our analysis that is important to signal at the outset: given the heterogeneity of footnote markers as well as their printed amiguity (footnotes can be designated by numerous different shapes which are very hard to distinguish from other marks or blemishes on the reproduced page), our analysis does not indicate where in the body of the text the footnote is anchored. In other words, we cannot provide analysis of the footnoted word, but only an estimation of the presence of the footnote itself at the bottom of the page. Further research would be needed to reliably capture the location of the footnote mark as indicated in Figure 1.

Based on the above definition and limitations, and with extensive discussions between students and faculty responsible for collecting training data, we manually annotated 21,939 page images for training (6,028 pages with footnotes and 15,911 pages without), and another 5,520 pages for testing (522 with and 4,998 without). All pages were randomly generated from ECCO I and II and then reviewed by a single student. Ambiguous cases were reviewed by the project investigators. As we will demonstrate, our models do not appear to show biases towards different historical timeframes within the overall dataset or between ECCO I and II, which are collected separately by Gale.


Footnote Mark
Main Text
Footnote Mark
Footer Text
Catchword (Not a footnote)

[16]

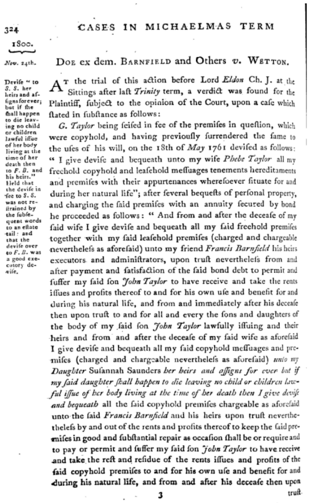
wealth: *ſo that it is not much to be wondered at, if, in the Progreſs of Years, publick Rapine, Violence, and Cruelty, Diſorders, and Licentiouſneſs of every Kind, were practiſed with Impunity; 'till Rome became as famous by her Debaucheries and Crimes, as by her Victories. It cannot however be denied, that, even in the moſt corrupted Times, there were not wanting ſome few bright Examples of Temperance, and true Patriotiſm; who will receive from lateſt Poſterity the Praiſes due to their uncommon Virtue; and will be reſpected as Ornaments of human Nature, while they are the Reproach of their Age and Country.

The Roman Conſtitution is highly magnified by ſome Writers, as being neither a Monarchy, nor an Ariſtocracy, nor yet a Popular Government; but ſuch a Mixture of all theſe, as united the Advantages, and avoided the Dangers, and

* Theſe Vices, joined to their repeated Breach of Faith, were the Cauſes of the firſt great Sedition that happened in Rome. The People retired from the City; and the Senate, to appeaſe them, fooliſhly conſented to the Creation of Tribunes: that is, they gave an incurable wound to their own Authority and Independency; and, in effect, changed the Conſtitution. Theſe Officers, it is well known, proved the Occaſion of perpetual Diſſenſions; and, which deſerves to be remarked, became at laſt the Inſtruments of ſubverting that very Liberty, which they had been appointed to defend.

Incon-

Figure 1. Example of a footnote. From Reflections on ancient and modern history (1742).



324 CASES IN MICHAELMAS TERM

1800.

Nov. 24th.

Doe ex dem. Barnfield and Others *v.* Wetton.

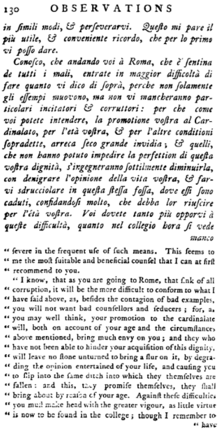
Deviſe " to *S. S.* her heirs and aſſigns for ever; but if ſhe ſhall happen to die leaving no child or children lawful iſſue of her body living at the time of her death then to *F. B.* and his heirs." Held that the deviſe in fee to *S. S.* was not reſtrained by the ſubſequent words to an eſtate tail: and that the deviſe over to *F. B.* was a good executory deviſe.

At the trial of this action before Lord *Eldon* Ch. J. at the Sittings after laſt *Trinity* term, a verdict was found for the Plaintiff, ſubject to the opinion of the Court, upon a caſe which ſtated in ſubſtance as follows:

G. Taylor being ſeiſed in fee of the premiſes in queſtion, which were copyhold, and having previouſly ſurrendered the ſame to the uſes of his will, on the 18th of *May* 1761 deviſed as follows: " I give deviſe and bequeath unto my wife *Phebe Taylor* all my freehold copyhold and leaſehold meſſuages tenements hereditaments and premiſes with their appurtenances whereſoever ſituate for and during her natural life"; after ſeveral bequeſts of perſonal property, and charging the ſaid premiſes with an annuity ſecured by bond he proceeded as follows: " And from and after the deceaſe of my ſaid wife I give deviſe and bequeath all my ſaid freehold premiſes together with my ſaid leaſehold premiſes (charged and chargeable nevertheleſs as aforeſaid) unto my friend *Francis Barnfield* his heirs executors and adminiſtrators, upon truſt nevertheleſs from and after payment and ſatisfaction of the ſaid bond debt to permit and ſuffer my ſaid ſon *John Taylor* to have receive and take the rents iſſues and profits thereof to and for his own uſe and benefit for and during his natural life, and from and immediately after his deceaſe then upon truſt to and for all and every the ſons and daughters of the body of my ſaid ſon *John Taylor* lawfully iſſuing and their heirs and from and after the deceaſe of my ſaid wife as aforeſaid I give deviſe and bequeath all my ſaid copyhold meſſuages and premiſes (charged and chargeable nevertheleſs as aforeſaid) *unto my Daughter* Suſannah Saunders *her heirs and aſſigns for ever but if my ſaid daughter ſhall happen to die leaving no child or children lawful iſſue of her body living at the time of her death then I give deviſe and bequeath* all the ſaid copyhold premiſes chargeable as aforeſaid unto the ſaid *Francis Barnfield* and his heirs upon truſt nevertheleſs by and out of the rents and profits thereof to keep the ſaid premiſes in good and ſubſtantial repair as occaſion ſhall be or require and to pay or permit and ſuffer my ſaid ſon *John Taylor* to have receive and take the reſt and reſidue of the rents iſſues and profits of the ſaid copyhold premiſes to and for his own uſe and benefit for and during his natural life, and from and after his deceaſe then upon

3 truſt

Fig. 2 Figure 2. Example of side notation. From Reports of cases argued and determined in the Courts of Common Pleas (1802).



130 OBSERVATIONS

in ſimili modi, & perſeverarvi. Queſto mi pare il più utile, & conveniente ricordo, che per lo primo vi poſſo dare.

Conoſco, che andando voi à Roma, che è ſentina de tutti i mali, entrate in maggior difficoltà di fare quanto vi dico di ſoprà, perche non ſolamente gli eſſempi muovono, ma non vi mancheranno particolari incitatori & corruttori: per che come voi potete intendere, la promotione voſtra al Cardinalato, per l'età voſtra, & per l'altre conditioni ſopradette, arreca ſeco grande invidia; & quelli, che non hanno potuto impedire la perfettion di queſta voſtra dignità, s'ingegneranno ſottilmente diminuirla, con denigrare l'opinione della vita voſtra, & farvi sdrucciolare in queſta ſteſſa foſſa, dove eſſi ſono caduti, confidandoſi molto, che debba lor riuſcire per l'étà voſtra. Voi dovete tanto più opporvi à queſte difficultà, quanto nel collegio hora ſi vede

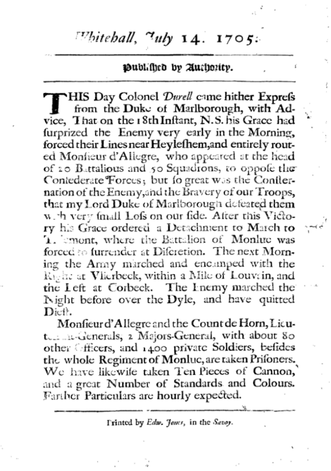
manco

" ſevere in the frequent uſe of ſuch means. This ſeems to " me the moſt ſuitable and beneficial counſel that I can at firſt " recommend to you.

" I know, that as you are going to Rome, that ſink of all " corruption, it will be the more difficult to conform to what I " have ſaid above, as, beſides the contagion of bad examples, " you will not want bad counſellors and ſeducers; for, as " you may well think, your promotion to the cardinalate " will, both on account of your age and the circumſtances " above mentioned, bring much envy on you; and they who " have not been able to hinder your acquiſition of this dignity, " will leave no ſtone unturned to bring a ſlur on it, by degra- " ding the opinion entertained of your life, and cauſing you " to ſlip into the ſame ditch into which they themſelves are " fallen: and this, they promiſe themſelves, they ſhall " bring about by reaſon of your age. Againſt theſe difficulties " you muſt make head with the greater vigour, as little virtue " is now to be found in the college; though I remember to

" have

Figure 3. Example of commentary at the bottom of the page. From New observations on Italy and its inhabitants (1769).



Whitehall, July 14. 1705.

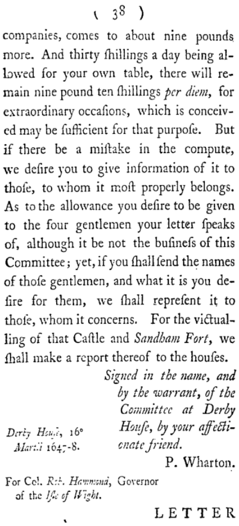
Publiſhed by Authority.

THIS Day Colonel *Durell* came hither Expreſs from the Duke of Marlborough, with Advice, That on the 18th Inſtant, N.S. his Grace had ſurprized the Enemy very early in the Morning, forced their Lines near Heyleſhem, and entirely routed Monſieur d'Allegre, who appeared at the head of 20 Battalions and 50 Squadrons, to oppoſe the Confederate Forces; but ſo great was the Conſternation of the Enemy, and the Bravery of our Troops, that my Lord Duke of Marlborough defeated them with very ſmall Loſs on our ſide. After this Victory his Grace ordered a Detachment to March to Tirlemont, where the Battalion of Monluc was forced to ſurrender at Diſcretion. The next Morning the Army marched and encamped with the Right at Vlierbeck, within a Mile of Louvain, and the Left at Corbeck. The Enemy marched the Night before over the Dyle, and have quitted Dieſt.

Monſieur d'Allegre and the Count de Horn, Lieutenant-Generals, 2 Majors-General, with about 80 other Officers, and 1400 private Soldiers, beſides the whole Regiment of Monluc, are taken Priſoners. We have likewiſe taken Ten Pieces of Cannon, and a great Number of Standards and Colours. Farther Particulars are hourly expected.

Printed by *Edw. Jones*, in the *Savoy*.

Figure 4. Example of footnote-like text in an early newspaper (1702).



(38)

companies, comes to about nine pounds more. And thirty ſhillings a day being allowed for your own table, there will remain nine pound ten ſhillings *per diem*, for extraordinary occaſions, which is conceived may be ſufficient for that purpoſe. But if there be a miſtake in the compute, we deſire you to give information of it to thoſe, to whom it moſt properly belongs. As to the allowance you deſire to be given to the four gentlemen your letter ſpeaks of, although it be not the buſineſs of this Committee; yet, if you ſhall ſend the names of thoſe gentlemen, and what it is you deſire for them, we ſhall repreſent it to thoſe, whom it concerns. For the victualling of that Caſtle and *Sandham Fort*, we ſhall make a report thereof to the houſes.

Signed in the name, and
by the warrant, of the
Committee at Derby
Houſe, by your affecti-
onate friend.

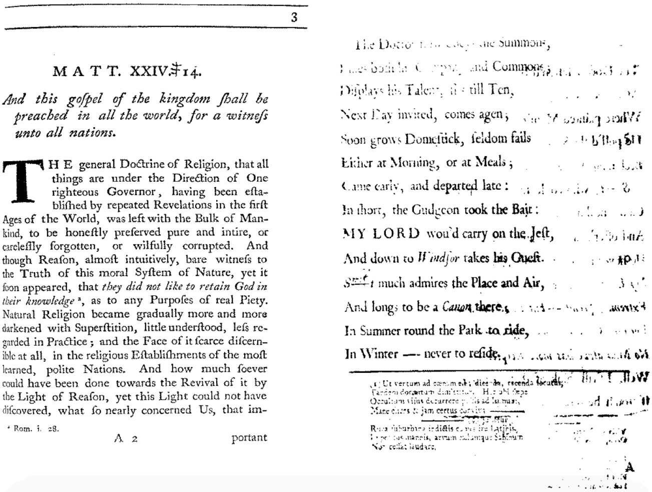
Derby Houſe, 16°
Martii 1647-8.

P. Wharton.

For Col. *Rob. Hammond*, Governor
of the *Iſle of Wight*.

LETTER

Figure 5. Example of footnote-like text. From Letters between Col. Robert Hammond (1764)



3

M A T T. XXIV. 14.

And this goſpel of the kingdom ſhall be preached in all the world, for a witneſs unto all nations.

THE general Doctrine of Religion, that all things are under the Direction of One righteous Governor, having been eſtabliſhed by repeated Revelations in the firſt Ages of the World, was left with the Bulk of Mankind, to be honeſtly preſerved pure and intire, or careleſſly forgotten, or wilfully corrupted. And though Reaſon, almoſt intuitively, bare witneſs to the Truth of this moral Syſtem of Nature, yet it ſoon appeared, that *they did not like to retain God in their knowledge* [a], as to any Purpoſes of real Piety. Natural Religion became gradually more and more darkened with Superſtition, little underſtood, leſs regarded in Practice; and the Face of it ſcarce diſcernible at all, in the religious Eſtabliſhments of the moſt learned, polite Nations. And how much ſoever could have been done towards the Revival of it by the Light of Reaſon, yet this Light could not have diſcovered, what ſo nearly concerned Us, that im-

[a] Rom. i. 28.

A 2 portant

The Doctor [illegible] the Summons,
[illegible] both in [illegible] and Commons;
Diſplays his Talent [illegible] till Ten,
Next Day invited, comes agen;
Soon grows Domeſtick, ſeldom fails
Either at Morning, or at Meals;
Came early, and departed late:
In ſhort, the Gudgeon took the Bait:
MY LORD wou'd carry on the Jeſt,
And down to *Windſor* takes his Gueſt.
S——*t* much admires the Place and Air,
And longs to be a *Canon* there;
In Summer round the Park to ride,
In Winter —— never to reſide.

(1) Ut ventum ad coenam eſt, dicenda, tacenda locutus,
Tandem dormitum dimittitur. Hic ubi ſæpe
Occultum viſus decurrere piſcis ad hamum,
Mane cliens & jam certus conviva ———
[illegible]
Rura ſuburbana indictis comes ire Latinis,
Impoſitus mannis, arvum cœlumque Sabinum
Non ceſſat laudare.

A

Figure 6. Examples of degraded or hard to capture footnote-marks. Part of the seventh epistle of the First book of Horace (1713) (left) and A sermon by Joseph Lord Bishop of Bristol (1739) (right).

Detecting Footnotes at Large Scale (Machine Learning)

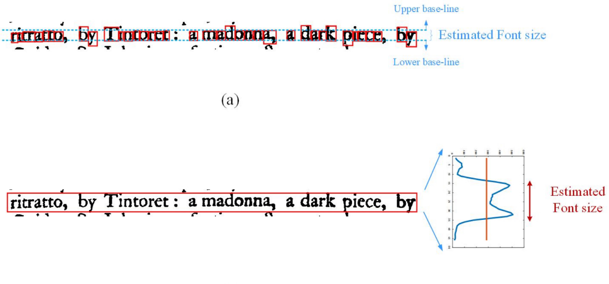
After collecting our positive and negative examples of footnoted pages, we then set out to design features and learning algorithms that could best predict the presence of footnotes on a page. We chose to use four models which we describe here. The performance of the models is reported in Table 1. The designed models cover a range of machine learning approaches, beginning with a conventional model based only on hand-designed features and moving to a learning-based model that utilizes deep learning.

Our first approach is a "rule-based" model that tries to capture three overarching visual features related to the differential line-size and line-spacing of pages (thus "rule-based"). Our hypotheses for this model are that footnotes will: 1) have a smaller font size than the main text; 2) be located at the bottom of the page; and 3) be indicated by significant spacing between the footnote and the main text. The advantage of this kind of approach is that the creation of custom features can target our prior knowledge of the problem (i.e. what is a footnote) and increase precision. The drawback is that the delimitation of features may not be able to capture the broader diversity of footnote behaviors in our data and thus may lower recall. This can be compensated for by more learning-based approaches where features are not pre-defined but learned from the training examples. However, as we show in Table 1, we see how overall in our results we do achieve higher precision (finding true positives) and lower recall (producing false negatives, i.e. overlooking footnotes).

In order to estimate font size (hypothesis 1), we use two methods drawn from the field of document image analysis: the bounding box method (BBox) and the horizontal projection method (Proj) (Fig. 7).[3] Bounding boxes are determined for each line by finding the rectangles containing the connected components. A connected component is defined as the continuous connection of black pixels. In theory, a connected component should correspond to individual letters, but given the imperfect reproductions of pages along with typographic irregularities

[3] Rodolfo P. dos Santos, et al., "Text line segmentation based on morphology and histogram projection," *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on.* IEEE, 2009 and Laurence Likforman-Sulem, Abderrazak Zahour, and Bruno Taconet, "Text line segmentation of historical documents: a survey," *International Journal of Document Analysis and Recognition (IJDAR)* 9.2-4 (2007): 123-138.

introduced in historical printing practices, errors can be introduced (Fig. 8). These bounding boxes are then used to estimate the lower case letters' font sizes by finding the distance between the lower and the upper base-lines, as shown in Fig. 7(a). For the horizontal projection method, the horizontal intensity for each line is calculated (i.e., the pixels in the horizontal direction are summed such that there will be fewer pixels at the upper and lower levels of the line where extenders and descenders are located (capital letters or d's or y's for example)). As demonstrated in Fig. 7(b), the font size of a textline is estimated by calculating the distance of the inner intersected line between the derived projection and a threshold line of a value equal to 0.55.


Upper base-line
ritratto, by Tintoret : a madonna, a dark piece, by
Estimated Font size
Lower base-line
ritratto, by Tintoret : a madonna, a dark piece, by
Estimated
Font size

(a)

(b)

Figure 7. The estimated font size of (a) the bounding box based method and (b) the horizontal projection based method.

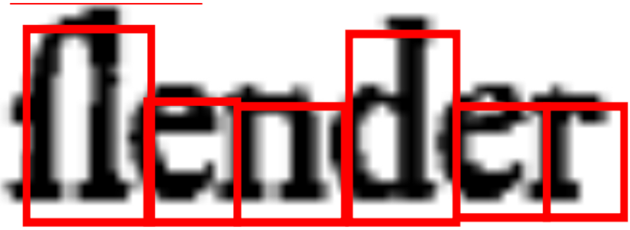


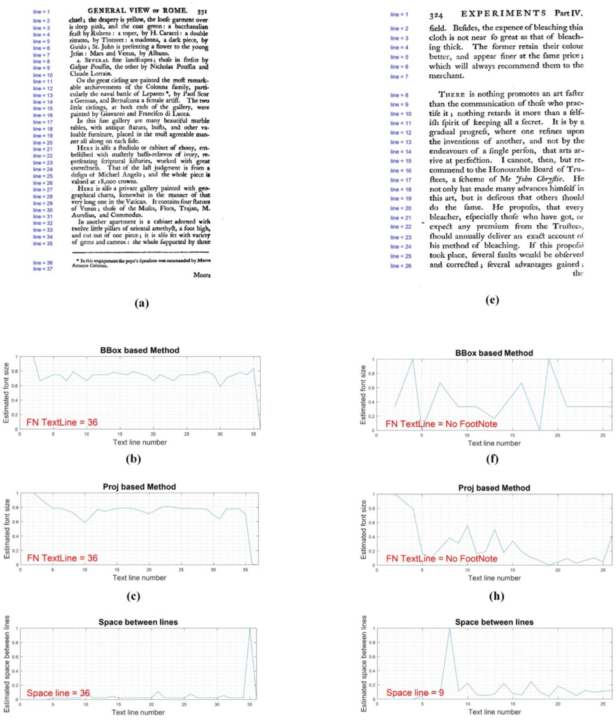
Figure 8. We can see in this example of the word "slender" from Hogarth's Analysis of Beauty a connected component that spans more than one letter due to the typeface used and the potential for bleeding between letters. Each red box represents a connected component.

According to our initial hypothesis, we expect that any line with a footnote would correspond to a decrease in font size when compared to the previous line. Although this would be an ideal case, such a decrease could be attributable to something other than the presence of a footnote, such as the presence of a title, figure, or tables, etc. We therefore define additional rules in order to improve the accuracy of our footnote detection. We convert these rules into specific features described in Appendix A.[4]

In order to identify the footnote location (our second assumption), we define a series of further rules based on the page layout for each method (BBox and Proj). The relative position of the estimated footnote line to all other lines on the document image is then used as a basis of further features. The third and final technique determines the spaces between the lines and uses them as a feature. More specifically, the textline below the large white space closest to the bottom of the page is considered to be a footnote candidate. The location of a possible footnote is compared to the locations estimated in the first two methods in order to partially form the final feature vector of the image. We illustrate how the three primary features of line height, line spacing, and page location perform with respect to two sample pages, one with a footnote, one without (Fig. 9).

Using these three primary features we develop a total of 72 features related to rule-based qialities of the page (18 BBox + 24 Proj + 30 Location and space as described in Appendix A) which are then fed into a support vector machine (SVM) classifier to detect pages with footnotes.

[4] Sara Zhalehpour, Andrew Piper, Chad Wellmon, and Mohamed Cheriet, "Footnote-based document image classification," In *International Conference Image Analysis and Recognition* (Cham: Springer, 2017): 634-642.



GENERAL VIEW of ROME. 331

chael; the drapery is yellow, the looſe garment over is deep pink, and the coat green: a bacchanalian feaſt by Rubens: a toper, by H. Caracci: a double ritratto, by Tintoret: a madonna, a dark piece, by Guido; St. John is preſenting a flower to the young Jeſus: Mars and Venus, by Albano.

4. Several fine landſcapes; thoſe in freſco by Gaſpar Pouſſin, the other by Nicholas Pouſſin and Claude Lorrain.

On the great cieling are painted the moſt remarkable atchievements of the Colonna family, particularly the naval battle of Lepanto *, by Paul Scor a German, and Bernaſcona a female artiſt. The two little cielings, at both ends of the gallery, were painted by Giovanni and Franciſco di Lucca.

In this fine gallery are many beautiful marble tables, with antique ſtatues, buſts, and other valuable furniture, placed in the moſt agreeable manner all along on each ſide.

Here is alſo a ſtudiolo or cabinet of ebony, embelliſhed with maſterly baſſo-relievos of ivory, repreſenting ſcriptural hiſtories, worked with great correctneſs. That of the laſt judgment is from a deſign of Michael Angelo; and the whole piece is valued at 18,000 crowns.

. Here is alſo a private gallery painted with geographical charts, ſomewhat in the manner of that very long one in the Vatican. It contains four ſtatues of Venus; thoſe of the Muſes, Flora, Trajan, M. Aurelius, and Commodus.

In another apartment is a cabinet adorned with twelve little pillars of oriental amethyſt, a foot high, and cut out of one piece; it is alſo ſet with variety of gems and cameos: the whole ſupported by three

* In this engagement the pope's ſquadron was commanded by Marco Antonio Colonna.

Moors

(a)

324 EXPERIMENTS Part IV.

field. Beſides, the expence of bleaching thin cloth is not near ſo great as that of bleaching thick. The former retain their colour better, and appear finer at the ſame price; which will always recommend them to the merchant.

There is nothing promotes an art faſter than the communication of thoſe who practiſe it; nothing retards it more than a ſelfiſh ſpirit of keeping all a ſecret. It is by a gradual progreſs, where one refines upon the inventions of another, and not by the endeavours of a ſingle perſon, that arts arrive at perfection. I cannot, then, but recommend to the Honourable Board of Truſtees, a ſcheme of Mr *John Chryſtie*. He not only has made many advances himſelf in this art, but is deſirous that others ſhould do the ſame. He propoſes, that every bleacher, eſpecially thoſe who have got, or expect any premium from the Truſtees, ſhould annually deliver an exact account of his method of bleaching. If this propoſal took place, ſeveral faults would be obſerved and corrected; ſeveral advantages gained;

the

(e)

(b) (f)

(c) (h)

(d) (i)

Figure 9. Examples of two document pages, with footnote (a) and without footnote (b), and their related results after applying the BBox-based method (b, f), the projection-based method (c, h) and estimating the lines spacing (d, i).

Our second model is a "layout-based" model that combines hand-designed and learned features, although it weighs the former more heavily. This and the next model might be considered to be hybrid models that combine custom features defined by expert knowledge and learned features defined by the machine's exposure to the training data. The layout approach primarily depends on understanding the layout of textlines on a page (Fig. 10). Once again, it rests on the hypothesis that footnotes will exhibit distinctive visual behavior with respect to their size and position on the page. Similar to the rule-based approach, we develop 22 custom measures for each textline based on the variables shown in Fig. 10 (see Appendix B for a full description of all features). Because the number of textlines varies between document images, it is necessary to extract features with a fixed length for all of the images. In order to do this, we use Discrete Cosine

Transform (DCT),[5] where we consider the concatenation of each textline's 22 measures as a signal. This signal exhibits a repetitive behavior and thus contains frequency information such that DCT can be used to capture this information. Specifically, since most of the signal's energy (i.e., information) is concentrated in lower frequencies, and assuming that document images have at least 5 textlines, we kept only the first 300 coefficients of the DCT transform for each image.

The final step of the "layout-based" model is classification. We use a combination of an Autoencoder overlaid with a softmax layer.[6] The Autoencoder creates lower dimensional representation of the provided input data in its hidden layer and then reconstructs this data at its output layer. This representation is then fed to a softmax layer with the labels of the document images to learn the model for classifying new samples.

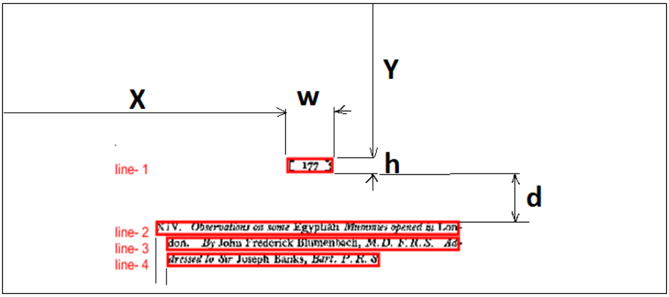

Y
X
w
h
d
line- 1
[177]
line- 2
line- 3
line- 4
XIV. Observations on some Egyptian Mummies opened in London. By John Frederick Blumenbach, M.D. F.R.S. Addressed to Sir Joseph Banks, Bart. P.R.S

Figure 10. A bounding box of a textline with features X, w, Y, h, and d defined as relative positions on the page.

Given the recent advances in the field of deep learning, particularly with architectures such as Convolutional Neural Networks (CNN), we used two CNN-based models for our final two approaches.[7] We also used two different techniques in an effort to compensate for the limited amount of labeled data available to us, because CNNs generally require large amounts of data during the training process in order to perform efficiently.

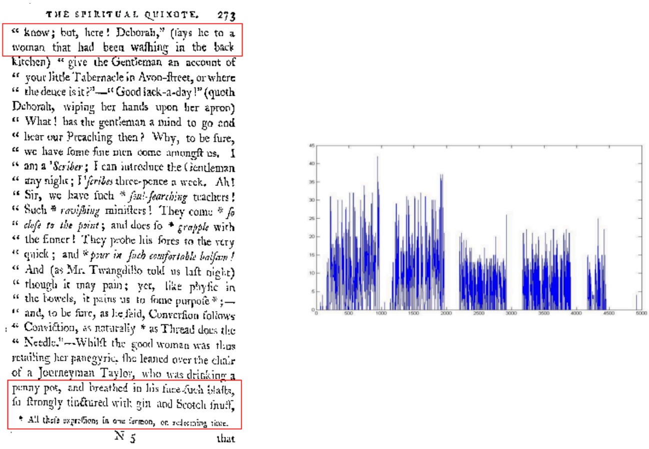
In the third approach, "CNN-based," the model is based primarily on learning the document image's features throughout the various layers of the neural network. But the model also depends on hand-designed features in order to overcome the limited amount of labeled data. Based on our hypothesis that the foot-

[5] Edmund Y. Lam, "Analysis of the DCT coefficient distributions for document coding," IEEE Signal Processing Letters 11.2 (2004): 97-100.

[6] See Pierre Baldi, "Autoencoders, unsupervised learning, and deep architectures," In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012.

[7] Ian Goodfellow, et al., *Deep learning*. Vol. 1. (Cambridge: MIT press, 2016).

note's text and the main body's text differ in both style and font, each document image is represented using the two top textlines and the three bottom ones. (We use a projection-based segmentation method described above to detect those textlines.) Each of these textlines is represented, more precisely, as a vertical histogram (Fig. 11). As in the previous examples, the performance of this model will be hindered by the reliance on layout assumptions that may not always apply to our object of study. In order to capture changes in font size, here we use vertical projections of the lines, meaning the bars of the histogram represent vertical slices of the red-bounded lines. The lower height of the histogram bars represents a lower average line-height. A concatenated version of the vertical histograms of these textlines is then used as an input to a 1-dimensional CNN (i.e., a 5000x1 histogram).[8]



THE SPIRITUAL QUIXOTE. 273

" know; but, here! Deborah," (ſays he to a
woman that had been waſhing in the back
kitchen) " give the Gentleman an account of
" your little Tabernacle in Avon-ſtreet, or where
" the deuce is it?"—" Good lack-a-day!" (quoth
Deborah, wiping her hands upon her apron)
" What! has the gentleman a mind to go and
" hear our Preaching then? Why, to be ſure,
" we have ſome fine men come amongſt us. I
" am a '*Scriber*; I can introduce the Gentleman
" any night; I '*ſcribes* three-pence a week. Ah!
" Sir, we have ſuch * *ſoul-ſearching* teachers!
" Such * *raviſhing* miniſters! They come * *ſo*
" *cloſe to the point*; and does ſo * *grapple* with
" the ſinner! They probe his ſores to the very
" quick; and * *pour in ſuch comfortable balſam!*
" And (as Mr. Twangdillo told us laſt night)
" though it may pain; yet, like phyſic in
" the bowels, it pains us to ſome purpoſe *;—
" and, to be ſure, as he ſaid, Converſion follows
" Conviction, as naturally * as Thread does the
" Needle."—Whilſt the good woman was thus
retailing her panegyric, ſhe leaned over the chair
of a Journeyman Taylor, who was drinking a
penny pot, and breathed in his face ſuch blaſts,
ſo ſtrongly tinctured with gin and Scotch ſnuff,

* All theſe expreſſions in one ſermon, on releaſing time.

N 5 that

(a) (b)

Figure 11. A representation of a document page (a) with its related vertical histograms (b).

Our fourth and final approach is based on transfer learning and CNN ("Transfer Learning").[9] According to this approach, the model automatically learns the features without using any hand-designed features. Transfer learning can be par-

[8] Mohamed Mhiri, Sherif Abuelwafa, Christian Desrosiers, and Mohamed Cheriet, "Footnote-based Document Image Classification using 1D Convolutional Neural Networks and Histograms," In *International Conference on Image Processing Theory, Tools and Applications*, 2017.

[9] Yoshua Bengio, Aaron Courville, and Pascal Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013): 1798-1828.

ticularly useful given the scarcity of labeled training data in our case. The model consists of two supervised learning stages, a pre-training stage and a fine-tuning stage. In the first stage (pre-training), an AlexNet is trained on a large dataset of natural images and the resultant learned parameters (e.g., network's weights) are saved. In the second stage (fine-tuning), instead of initializing the CNN's parameters randomly, the model uses learned parameters from stage one. Then, we use the ECCO dataset to train the model to classify document images with footnotes. The novelty of this approach is important to emphasize footnotes are being learned first by learning features of "images" more generally and then being trained on page images more specifically. To prepare the data, we perform two pre-processing steps on the raw document images—resizing and normalization—before using them as inputs to our model. Each document image is resized to 227x227, and its pixel values are normalized to be in the range [0 1] (Fig. 12). Unlike the first three approaches, this model does not require any textline segmentation process; therefore, it avoids the segmentation errors that may result from it.

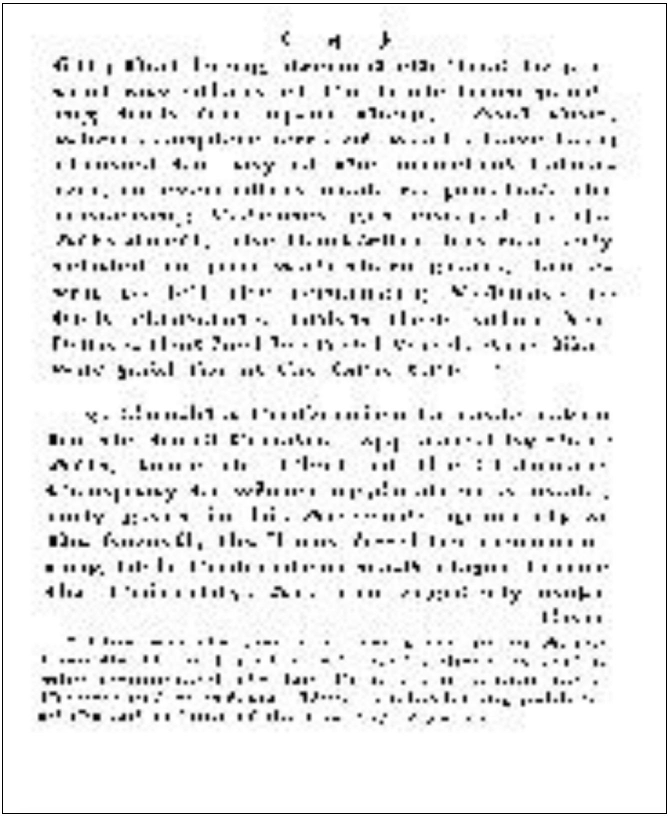


Figure 12. In this example, pages are binarized and then reduced in size to 227x227 pixels (or 51,529 dimensions) rendering them illegible, but ideally capturing the unique visual signature of footnotes.

These then were the four models we developed to detect footnotes. As a final step, we use an ensemble detection method that combines all four classifiers. Applying this ensemble method on the test set of ECCO, we achieve 96.2% precision and 67.87% recall in our footnote detection results (Table 1).

Approaches	Precision (%)	Recall (%)	F1 score
Rule-based	68.24%	60.8%	0.643
Layout-based	60.8%	69.4% 0.	6482
CNN-based	90.35%	48.37%	0.63
Transfer learning-based	74.31%	41.49%	0.5325
The final detection approach	**96.2%**	**67.87%**	**0.7959**

Table 1. The individual performance of each detection approach, in addition to

the final approach performance.

Applying our detection methods on the full ECCO dataset, we discovered 1,319,000 footnoted images from approximately 26,000,000 document image in ECCO I and 239,754 footnoted images from approximately 6,000,000 document image in ECCO II. We therefore estimate that roughly 4.9% of all pages in the eighteenth contained footnotes. The figures below (Figs. 13-15) provide more detailed results, including the number of document images with detected footnotes over time (publication years) as well as document images with detected footnotes according to subject classifications in both ECCO I and II. We expect in a separate piece to explore this data in more detail. We share the underlying metadata of footnote annotation to allow others to do the same.

We also provide detailed information in Tables 2-3 that demonstrate the consistency of our final model's performance across different time periods and subjects in both ECCO I and II. As we show, the values of the average footnote probability per page are stable (i.e., around 0.68) regardless of the year or subject of the examined document image. These tables give us confidence that our predicted levels of footnotes are not dependent on either document type or the year of publication. All of our derived data has been shared as supplementary data to this article.

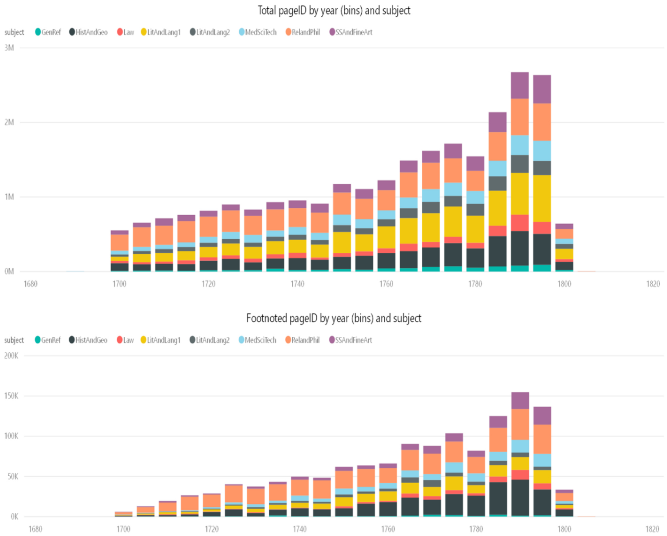


Figure 13. Distribution of document images (all and footnoted) in ECCO I by year using Gale's eight subject classes.

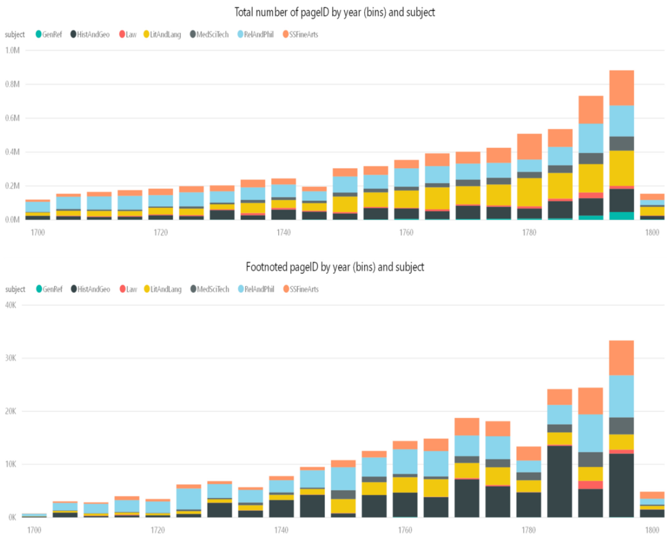


Figure 14. Distribution of document images (all and footnoted) in ECCO II by year using Gale's eight subject classes.

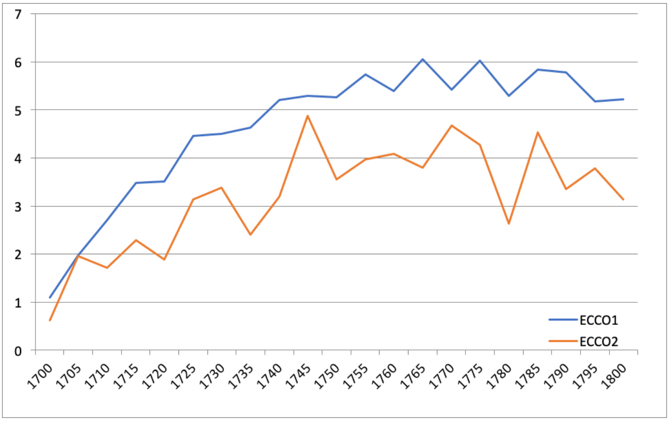
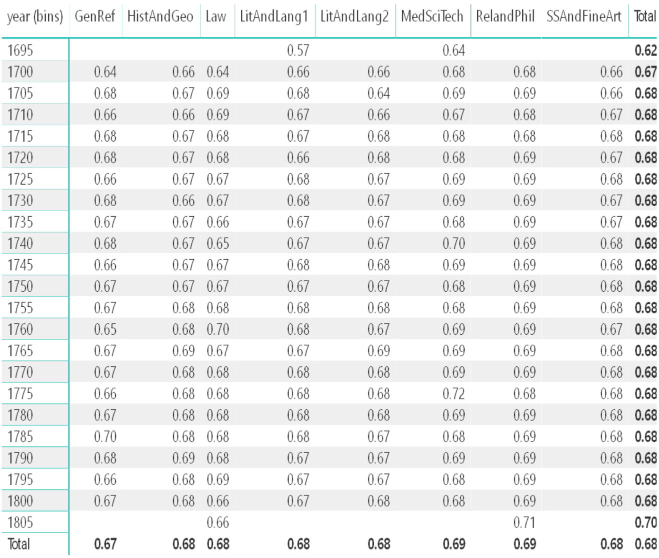
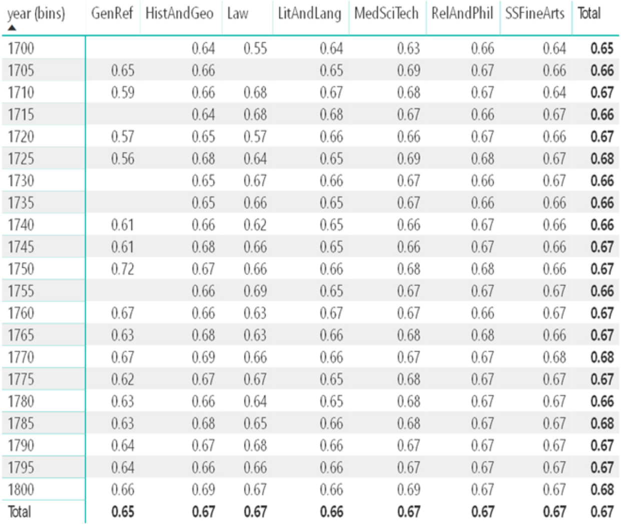


Figure 15. The percentage (%) of the detected footnote document images to the total document images at both ECCO I and ECCO II.



year (bins)	GenRef	HistAndGeo	Law	LitAndLang1	LitAndLang2	MedSciTech	RelandPhil	SSAndFineArt	Total
1695				0.57		0.64			**0.62**
1700	0.64	0.66	0.64	0.66	0.66	0.68	0.68	0.66	**0.67**
1705	0.68	0.67	0.69	0.68	0.64	0.69	0.69	0.66	**0.68**
1710	0.66	0.66	0.69	0.67	0.66	0.67	0.68	0.67	**0.68**
1715	0.68	0.67	0.68	0.67	0.68	0.68	0.68	0.68	**0.68**
1720	0.68	0.67	0.68	0.66	0.68	0.68	0.69	0.67	**0.68**
1725	0.66	0.67	0.67	0.68	0.67	0.69	0.69	0.68	**0.68**
1730	0.68	0.66	0.67	0.68	0.67	0.69	0.69	0.67	**0.68**
1735	0.67	0.67	0.66	0.67	0.67	0.68	0.69	0.67	**0.68**
1740	0.68	0.67	0.65	0.67	0.67	0.70	0.69	0.68	**0.68**
1745	0.66	0.67	0.67	0.68	0.68	0.69	0.69	0.68	**0.68**
1750	0.67	0.67	0.67	0.67	0.67	0.68	0.69	0.68	**0.68**
1755	0.67	0.68	0.68	0.68	0.68	0.68	0.69	0.68	**0.68**
1760	0.65	0.68	0.70	0.68	0.67	0.69	0.69	0.67	**0.68**
1765	0.67	0.69	0.67	0.67	0.69	0.69	0.69	0.68	**0.68**
1770	0.67	0.68	0.68	0.68	0.68	0.69	0.69	0.68	**0.68**
1775	0.66	0.68	0.68	0.68	0.68	0.72	0.68	0.68	**0.68**
1780	0.67	0.68	0.68	0.68	0.68	0.69	0.69	0.68	**0.68**
1785	0.70	0.68	0.68	0.68	0.67	0.68	0.69	0.68	**0.68**
1790	0.68	0.69	0.68	0.67	0.67	0.69	0.69	0.68	**0.68**
1795	0.66	0.68	0.69	0.67	0.67	0.68	0.69	0.68	**0.68**
1800	0.67	0.68	0.66	0.67	0.68	0.68	0.69	0.68	**0.68**
1805			0.66				0.71		**0.70**
Total	**0.67**	**0.68**	**0.68**	**0.68**	**0.68**	**0.69**	**0.69**	**0.68**	**0.68**

Table 2. The average probability of footnote by years and subjects (where images are detected as footnote) at ECCO I.



year (bins) ▲	GenRef	HistAndGeo	Law	LitAndLang	MedSciTech	RelAndPhil	SSFineArts	Total
1700		0.64	0.55	0.64	0.63	0.66	0.64	**0.65**
1705	0.65	0.66		0.65	0.69	0.67	0.66	**0.66**
1710	0.59	0.66	0.68	0.67	0.68	0.67	0.64	**0.67**
1715		0.64	0.68	0.68	0.67	0.66	0.67	**0.66**
1720	0.57	0.65	0.57	0.66	0.66	0.67	0.66	**0.67**
1725	0.56	0.68	0.64	0.65	0.69	0.68	0.67	**0.68**
1730		0.65	0.67	0.66	0.67	0.66	0.67	**0.66**
1735		0.65	0.66	0.65	0.67	0.66	0.66	**0.66**
1740	0.61	0.66	0.62	0.65	0.66	0.67	0.66	**0.66**
1745	0.61	0.68	0.66	0.65	0.66	0.67	0.66	**0.67**
1750	0.72	0.67	0.66	0.66	0.68	0.68	0.66	**0.67**
1755		0.66	0.69	0.65	0.67	0.67	0.67	**0.66**
1760	0.67	0.66	0.63	0.67	0.67	0.66	0.67	**0.67**
1765	0.63	0.68	0.63	0.66	0.68	0.68	0.66	**0.67**
1770	0.67	0.69	0.66	0.66	0.67	0.67	0.68	**0.68**
1775	0.62	0.67	0.67	0.65	0.68	0.67	0.67	**0.67**
1780	0.63	0.66	0.64	0.65	0.68	0.67	0.67	**0.66**
1785	0.63	0.68	0.65	0.66	0.68	0.67	0.67	**0.68**
1790	0.64	0.67	0.68	0.66	0.67	0.67	0.67	**0.67**
1795	0.64	0.66	0.66	0.66	0.67	0.67	0.67	**0.67**
1800	0.66	0.69	0.67	0.66	0.69	0.67	0.67	**0.68**
Total	**0.65**	**0.67**	**0.67**	**0.66**	**0.67**	**0.67**	**0.67**	**0.67**

Table 3. the average probability of footnote by years and subjects at ECCO II.

Appendix A: The Rule-based Footnote Detection Approach Features

The final utilized feature vector at the rule-based footnote detection approach contains 72 features, which is the combination of the features extracted at the following three techniques.

1) The Bounding Box (BBox) based Method:
At this method, 18 features are being utilized based on some initial assumptions; for instance, the assumption that the font size of any footnote line is at least 0.55 smaller than the font size of the main text. More assumptions are considered and demonstrated in details in table A1.

Table A1

Feature	Condition
1	1 if there is no drop more than 0.55
2	1 if there are 1+ drops of more than 0.55
3	1 if the last two lines' heights are less than 0.1
4	1 if the last lines' height is less than 0.1
5	1 if the last two lines' heights are less than 0.1 and there is a footnote
6	1 if the line before last line's height is less than 0.1
7	1 if the footnote is not in the 4th line
8	1 if there are 2+ drops more than 0.55
9	1 if the footnote is not in the 4th, 5th and 6th lines
10	1 if there are 2+ drops less than 0.55 or the footnote is not in the 4th, 5th and 6th lines
11	1 if there is a drop greater than 0.15
12	1 if footnote line is in the 6th line or later
13	1 if the height of the footnote line is 0.55 greater than the line before the last line
14	1 if there is a drop of greater than 0.35 between the lines before and after the footnote line
15	1 if there is a line except the last line selected as the footnote line and there is a drop of greater than 0.35 between the lines before and after it
16	1 if there is a difference less than 0.17 between the lines before and after the footnote line
17	1 if there is a line except the last line selected as the footnote line and there is a difference less than 0.17 between the lines before and after it
18	1 if it the page has more than 3 lines

2) The Horizontal Projection (Proj) based Method:
At this method, 24 features are being utilized. Table A2 demonstrates the extracted features in more details.

Table A2

Feature	Condition
1	1 if there are more than 3 lines in the page
2	1 if there is no possible footnote
3	1 if there is more than one possible footnote (drops with the amount of 0.55 or more)
4	1 if the footnote line is in the first 3 lines or there are more than 3 possible footnotes
5	1 if there are more than 3 possible footnotes or there are lines shorter than 0.13 but not footnote lines
6	1 if the footnote line is in the first 3 lines or there are lines shorter than 0.13 but not footnote lines
7	1 if the footnote line is in the first 3 lines

Feature	Condition
8	1 if there are more than 3 possible footnotes
9	1 if there are lines shorter than 0.13 but not footnote lines
10	1 if the footnote line is in the first 3 lines, there are 3+ possible footnotes or there are lines shorter than 0.13, but not footnote lines
11	1 if the last line or the line before it has a height less than 0.1
12	1 if the last line or the line before it has height less than 0.1 and there is a footnote
13	1 if the last line has a height less than 0.1
14	1 if the line before the last line has a height less than 0.1
15	1 if there still exists a footnote line
16	1 if the height of the last line is less than 0.4
17	1 if the last line has a height less than 0.1 and there exist a footnote line
18	1 if the line before the last line has a height less than 0.1 and there exist a footnote line
19	1 if the height of the last line is less than 0.4 and there is a footnote line and the last line or the line before has a height less than 0.1
20	1 if the greatest height drop is equal or greater than 0.4 and there is at least a 0.25 drop between the line before and after footnote
21	1 if there is at least a 0.25 drop between the line before and after the footnote and the footnote line's height is less than 0.4
22	1 if the greatest height drop is equal or greater than 0.4 and the height of the last line is less than 0.4
23	1 if the height of the last line is less than 0.4 and the greatest height drop is equal or greater than 0.4 and there is at least a 0.25 drop between the line before and after footnote
24	1 if the height of the footnote line is 0.4 below the highest height of all the other lines except the first 3 and last lines

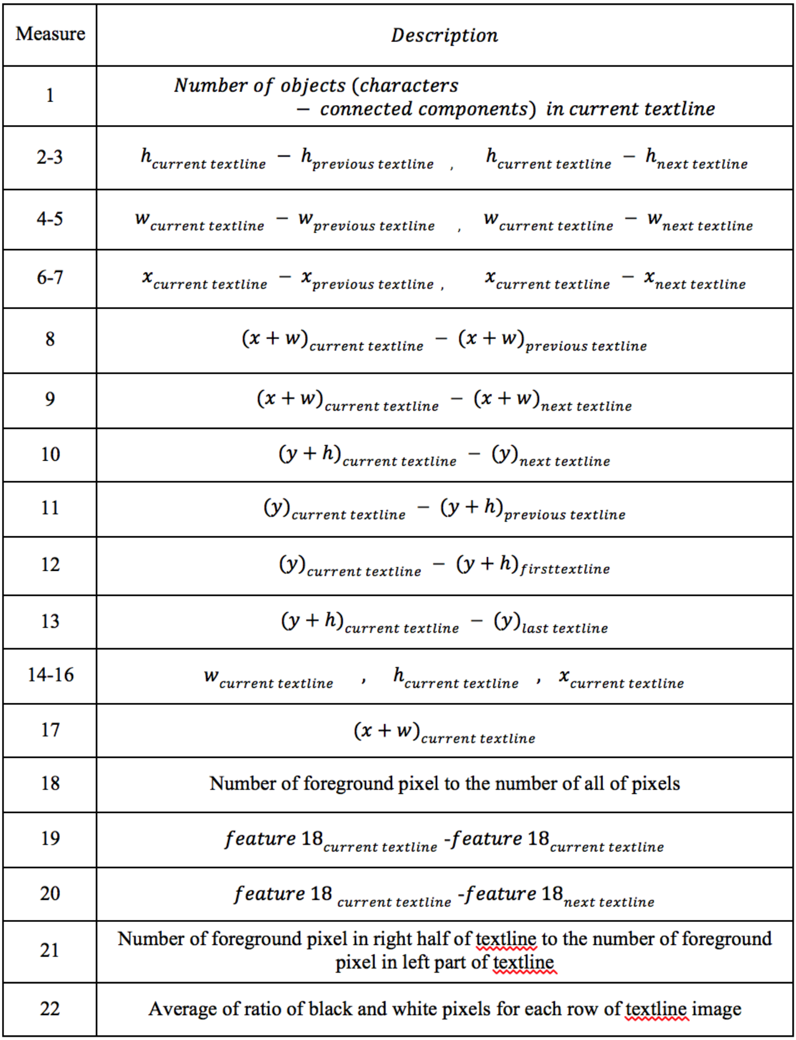
3) Location and Space based Features:
Table A3 demonstrates the 30 extracted features in more details.

Table A3

Feature	Condition
1	1 if there is more than 10 lines in the page
2	1 if there is a space peak in the 2nd 1/4th of the page and there is more than 10 lines in the page
3	1 if there is a space peak in the 3rd 1/4th of the page and there is more than 10 lines in the page
4	1 if there is a space peak in the 4th 1/4th of the page and there is more than 10 lines in the page
5	1 if there is only one peak in the page, select its location: (Peak location/ # of lines) and there is more than 10 lines in the page
6	1 if there is more than one peak in the page, select the last one?s location: (Peak location/ # of lines) and there is more than 10 lines in the page
"7-8"	1 if there is a footnote in the last 1/4th of the page: (FN location/ # of lines) and there is more than 10 lines in the page
"9-10"	1 if there is a footnote in the page: (FN location/ # of lines) and there is more than 10 lines in the page
"11-16"	Check if feature 5 appears anywhere around feature 7(Proj) using a threshold from ±0.02 by a 0.02 step and up to ±0.14
17-22	Check if feature 6 appears anywhere around feature 8(Proj) using a threshold from ±0.02 by a 0.02 step and up to ±0.14
23-26	Check if feature 5 appears anywhere around feature 7(BBox) using a threshold from ±0.02 by a 0.02 step and up to ±0.14
27-30	Check if feature 6 appears anywhere around feature 8(BBox) using a threshold from ±0.02 by a 0.02 step and up to ±0.14

Appendix B: The Layout-based Footnote Detection Approach Measures

A detailed description of the used measures at the layout-based footnote detection method is demonstrated at Table. B1



Measure	*Description*
1	$Number\ of\ objects\ (characters - connected\ components)\ in\ current\ textline$
2-3	$h_{current\ textline} - h_{previous\ textline}$, $h_{current\ textline} - h_{next\ textline}$
4-5	$w_{current\ textline} - w_{previous\ textline}$, $w_{current\ textline} - w_{next\ textline}$
6-7	$x_{current\ textline} - x_{previous\ textline}$, $x_{current\ textline} - x_{next\ textline}$
8	$(x+w)_{current\ textline} - (x+w)_{previous\ textline}$
9	$(x+w)_{current\ textline} - (x+w)_{next\ textline}$
10	$(y+h)_{current\ textline} - (y)_{next\ textline}$
11	$(y)_{current\ textline} - (y+h)_{previous\ textline}$
12	$(y)_{current\ textline} - (y+h)_{firsttextline}$
13	$(y+h)_{current\ textline} - (y)_{last\ textline}$
14-16	$w_{current\ textline}$, $h_{current\ textline}$, $x_{current\ textline}$
17	$(x+w)_{current\ textline}$
18	Number of foreground pixel to the number of all of pixels
19	$feature\ 18_{current\ textline}$ -$feature\ 18_{current\ textline}$
20	$feature\ 18_{current\ textline}$ -$feature\ 18_{next\ textline}$
21	Number of foreground pixel in right half of textline to the number of foreground pixel in left part of textline
22	Average of ratio of black and white pixels for each row of textline image

Table B1