

Visual Genealogy of Typography: Computational Provenance-Based Analysis of Typeface Evolution

Hendi Thamrin^{1*}, Shierly Everlin^{2*}, Puguh Hiskiawan^{3**},
Valentina^{4*}, Angelyn Stevanie Lee^{5*}, Paskalis Peter^{6**}

* Visual Communication Design, Faculty of Technology and Design, Bunda Mulia University

** Data Science, Faculty of Technology and Design, Bunda Mulia University

hthamrin@bundamulia.ac.id¹, severlin@bundamulia.ac.id², phiskiawan@bundamulia.ac.id³,
s34230053@student.ubm.ac.id⁴, s34230064@student.ubm.ac.id⁵, s36230020@student.ubm.ac.id⁶

Article Info

Article history:

Received 2026-09-24

Revised 2026-10-14

Accepted 2026-10-17

Keyword:

*Computational Typography,
Corpus Composition,
Morphological Similarity,
Typographic Provenance,
Visual Genealogy.*

ABSTRACT

Computational analyses of typographic corpora increasingly produce similarity networks that are read as evidence of stylistic descent, yet no study has tested what such morphology-based genealogical inference can legitimately support; this study provides that test. Using a quantitative re-analysis design, we took an existing dataset of 33 typeface families, with 1,188 glyph images and eleven morphological descriptors, and added an auditable provenance layer recording designer, foundry, dates, revival status and script category. Five diagnostics were applied at family level: chronological correspondence, script-category composition tested with Fisher's exact test and the adjusted Rand index, multiple-comparison correction, classification stability, and a comparison of documented derivations with undocumented pairs. Eleven of sixteen families designated historical carry design dates after the declared period of 1450–1899, and no rendered file belongs to it. The era groups differ in serif composition before any morphology is measured (50.0% versus 5.9%, $p = 0.0066$); recovered clusters align with script category rather than period (adjusted Rand index 0.534 versus 0.090) under all corpus configurations examined; and among sans-serif families no era difference is detected on the descriptors testable (all $p > 0.70$). Only two of eleven descriptor differences survive family-wise correction, a headline accuracy of 75.8% falls to a mean near 0.68 under repeated validation against a 51.5% baseline, and undocumented pairs reach similarities above documented derivations. Morphological similarity therefore measures formal affinity conditioned on corpus composition, a necessary but not sufficient basis for claims of descent. A five-point reporting protocol for computational typography is proposed.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Computational methods developed for textual corpora have been extended over the past decade to visual materials, producing a body of work concerned with what can and cannot be inferred from measurements of images [1, 2]. Typography is an unusually tractable candidate for this extension. Letterforms are discrete, standardised and machine-renderable; a typeface can be reproduced at arbitrary resolution under controlled conditions, and its constituent

glyphs isolated without segmentation error. Where most visual corpora require the analyst to decide what counts as an object, a typographic corpus arrives pre-segmented.

This tractability has supported a growing literature applying machine learning to letterforms. The dominant strands are optical character recognition, generative font synthesis, and font recognition and style classification for retrieval. Learned font manifolds and stroke-based representations model the space of typeface forms for generation and editing, while crowdsourced attribute models

learn a perceptual similarity metric that organises fonts for selection. In each case similarity is operational: it serves generation, navigation or retrieval, and carries no claim about historical relation. A smaller body of work treats formal properties of glyphs as objects of analysis, quantifying stroke, proportion and structure to support classification or comparison [3].

These developments raise a question the literature has not addressed directly. When morphological similarity between two typefaces is high, the finding is a statement about form. When that finding is presented as evidence that one typeface descends from another, the statement has changed category: it now asserts a historical relationship. The two are not equivalent. Two designs may converge on similar forms through independent response to a shared constraint screen rendering, industrial legibility requirements, geometric construction with no transmission between them. Convergence is not descent. Studies of material-culture evolution make the same distinction explicit: resemblance is admissible as evidence of descent only when inherited similarity can be separated from convergence and horizontal borrowing.

Type history already possesses the evidentiary apparatus that would adjudicate such claims. Designers, foundries, patents, specimen books and revival commissions record who drew what, when, and after which model. That record is not merely supplementary: it has repeatedly corrected genealogies that formal resemblance alone had sustained. The types issued as “Garamond” by several twentieth-century foundries were shown in 1926 to derive from Jean Jannon rather than Claude Garamond, an attribution error that survived three centuries of unimpeded access to the letterforms themselves [4]. Formal analysis did not detect it; provenance research did.

Computational typography has largely proceeded without engaging this record. A survey of Scopus-indexed studies applying machine learning or computer vision to typography since 2020 returns 204 records, of which one addresses ancestry or lineage, and does so as a stated motivation rather than as an analysis. The query combined *typograph**, *typeface* and *letterform* with machine learning, deep learning, computer vision and neural network terms, restricted to 2020 onward, articles and conference papers, English, retrieved 7 September 2026; 46 of the 204 records concern spelling correction or typosquatting rather than typography, giving an effective denominator of 158. Across the same set, only three mention serif classification at all, indicating that the script-category composition of analysed corpora is almost never reported.

The research gap is therefore not an absence of similarity measures, which are well developed, but the absence of any test of whether morphological similarity can support the genealogical claims attached to it. Existing studies either construct similarity for operational purposes, or detect glyph features without inferring relationships between typefaces [3]; none examines how an inference of descent is affected by

the chronology of the rendered artefacts, the composition of the corpus, the unit of analysis, or the stability of the reported separability. The limitation is general: a lineage inferred from form alone cannot distinguish transmission from convergence unless independent evidence of transmission is available.

A. Computational Analysis of Visual Artefacts

Arnold and Tilton [1] propose distant viewing as a framework for analysing large image collections, arguing that computer vision does not deliver neutral description but applies an interpretive code that must itself be made explicit. The framework has been extended into what Tilton and Wigard [2] term computational formalism: the analysis of visual media through measurable formal properties, with an accompanying account of what those measurements exclude. Two commitments recur and both are relevant here. Feature extraction encodes a theory of what matters in an image; a descriptor set is an argument, not an instrument. And computational output requires interpretive framing to become evidence, since a similarity score is a statement about representation rather than about the world.

Parallel developments in applied computer vision have established that convolutional architectures recover structured information from images where the relevant features are geometric rather than semantic. Hiskiawan et al. [5] demonstrate this for subsurface resistivity imaging, where convolutional networks extract structural patterns that conventional inversion methods represent only indirectly. Transferability to letterform analysis rests on the same premise: that glyph structure is a geometric property recoverable from a rendered image.

B. Typographic Analysis: From Principle-Based to Measurable

Typographic scholarship has traditionally proceeded through principle-based qualitative analysis. Erlyana and Hansen [6] exemplify this in an examination of Indonesian horror film posters, reading typographic choices against the principles of legibility, readability, visibility and clarity. The method is interpretive and expert-driven, and it produces knowledge computational analysis cannot: an account of why a form was chosen, situated in a production context.

A second body of work treats typography as a design variable within larger visual systems. Erlyana et al. [7] examine typographic layout as a designed artefact through a design-thinking process, while Everlin and Erlyana [8] trace the redesign of a corporate mark across successive versions, comparing formal attributes before and after revision a study of visual change conducted through qualitative comparison of documented cases. Semiotic approaches extend this to questions of value and meaning [9].

More recently, design research in this tradition has begun to quantify visual attributes directly. Erlyana et al. [10] operationalise colour as a measurable palette variable linked to affective response, and related work treats visual aesthetics as a construct amenable to statistical modelling rather than

expert judgement alone [11, 12]. The shift moves the unit of analysis from the interpreted example to the measured attribute, making design scholarship commensurable with computational method.

Computational work on typeface form has developed outside this tradition. Mason et al. [3] present the closest system to the present concern, detecting glyph anatomical features with convolutional networks and noting that such features carry information about a glyph's applications, ancestry and historical context. The system detects features; it does not infer relationships between typefaces.

C. Revival, Reissue, and Digital Re-Cutting

Typographic scholarship distinguishes carefully between a historical typeface and a contemporary artefact bearing its name. A typeface designed for metal type in the sixteenth century and a digital font file distributed under the same name in the twenty-first are different objects, produced under different technical constraints, by different designers, for different rendering conditions. The digital file is an interpretation of a historical model: proportions are regularised, optical sizes collapsed, spacing rebuilt for digital composition, and character sets extended far beyond the original.

The field's most instructive case concerns the name most likely to appear in any historical typeface corpus. The types issued as “Garamond” by several early twentieth-century foundries were based on matrices held at the Imprimerie Nationale and long attributed to Claude Garamond. In 1926 Beatrice Warde, writing as Paul Beaujon, demonstrated that these matrices were the work of Jean Jannon, cut around 1615 some seventy years after Garamond's death [4]. The correction is recounted in the standard reference literature on type history and practice [13, 14]. A substantial family of twentieth-century Garamonds therefore descends not from Garamond but from Jannon, while other digital cuts return to genuine Garamond sources. Two fonts sharing the name may have different ancestors, and a corpus entry labelled “Garamond, c. 1540” does not identify which.

The lesson is about method. The correction came from provenance research matrices, specimen books, archival records not from formal analysis of the letterforms, which had been available for inspection throughout. Form alone had sustained an incorrect genealogy for three centuries. This is precisely the limitation a morphology-based inference inherits. Classification systems have proved unable to absorb these distinctions. The Vox-ATypI system, devised in 1954 and adopted by the Association Typographique Internationale in 1962, organises typefaces by formal characteristics associated with particular centuries. It was formally de-adopted by ATypI in March 2021 on the grounds that it centres Latin script and no longer represents the field it was meant to describe [15]. No successor has been established. Classification is itself contested, which means a computational system cannot treat any existing scheme as ground truth and equally cannot proceed without one.

D. Corpus Composition and Methodological Reflexivity

Drucker [16] argues that data in the humanities is more accurately understood as *capta*, taken, not given and that visualisations inherit the interpretive decisions embedded in their construction. Applied to corpus design, the composition of a dataset is a finding-shaping variable rather than a preliminary detail. Where a corpus is divided into groups for comparison, any attribute differing systematically between those groups is a candidate confound. If the groups also differ in some property correlated with the measured features, an observed difference cannot be attributed to the grouping variable alone.

Related concerns apply to inference from small samples, where cross-validated estimates carry wide confidence intervals and vary substantially across resampling, and to repeated testing, where the family-wise error rate rises with the number of tests and correction procedures such as false discovery rate control [17] are required. Work on trustworthy computational frameworks has begun to formalise these requirements in applied settings [18, 19]. The present study takes a similar position for computational typography: provenance documentation, composition reporting and stability analysis belong to the method, not to the limitations paragraph.

E. Aim and Research Questions

The present study examines how far morphological similarity can support a claim of genealogical descent. Its focus is not the construction of a similarity network, which an earlier analysis of the same corpus within the authors' research group has already produced, but the evidentiary status of that network. It does not re-derive the morphological analysis. Instead it constructs an auditable provenance layer documenting designer, foundry, model and artefact dates, revival status and script category for each family in a typeface corpus, and re-reads an inferred genealogy against that layer. Four questions follow. The first (RQ1) asks how far the division between historical and contemporary families corresponds to the documented chronology of the artefacts actually analysed. The second (RQ2) asks whether the two groups differ systematically in script-category composition, and if so whether that composition accounts for the structure the analysis recovers. The third (RQ3) asks how stable the morphological separability between them proves under multiple-comparison correction and repeated validation. The fourth (RQ4) asks whether pairs of typefaces with a documented relation of derivation are distinguishable, in morphological similarity, from pairs without one.

II. METHOD

A. Corpus

The analysis draws on a typeface morphology dataset developed by the author's research group. The corpus comprises 33 typeface families, divided at construction into 16 designated historical (declared period 1450–1899) and 17

designated contemporary (2022–2026). Each family is represented by 36 standardised glyphs lowercase characters (a–z) and numerals (0–9) yielding 1,188 glyph images. Glyphs were rendered under a uniform pipeline: greyscale conversion, normalisation to a 1000×1000 pixel canvas, centre alignment and consistent rasterisation. Eleven morphological descriptors were extracted per glyph: shape compactness, stroke density, stroke contrast, character openness, whitespace ratio, character symmetry, aspect ratio, curvature, stroke direction, character entropy and visual complexity. Family-level values are the mean across that family's 36 glyphs. The present study takes this corpus as given and adds analytical layers to it; no glyphs were re-rendered and no descriptors recomputed. Where a value reported here differs from the parent study, the source of each value is stated.

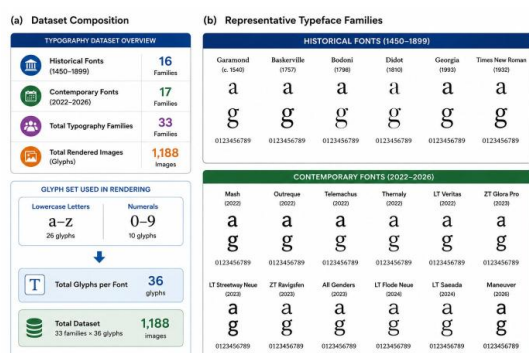


Fig. 1. Rendered specimens of representative families in the corpus.

B. Provenance Layer

A provenance layer was compiled for each of the 33 families from published type histories, foundry documentation and font file metadata, recording: the name as used in the corpus; the designer of the digital artefact; the year of the historical model where one exists; the year of release of the digital artefact; provenance status; and origin technology. Provenance status distinguishes four conditions. Original denotes a design with no prior model. Digitisation denotes transfer to digital outline with minimal reinterpretation. Revival denotes a new drawing based on a historical model by a different designer. Interpretation denotes a design drawing on a historical style without a specific model. Where documentation is insufficient the field is coded undocumented, treated as a finding rather than as missing data, since the availability of provenance is itself a property of the corpus. All entries are sourced to published documentation and are reproduced in full in Table I, so that each assignment can be independently checked.

C. Morphological Descriptor

Table II lists the eleven descriptors, the property each quantifies, and its typographic relevance. All descriptors are computed on the binarised glyph image at the common 1000×1000 resolution and are scale-free, so that values are

comparable across families; exact computational formulations follow the parent study.

Before clustering, classification and similarity computation, family-level descriptors were standardised across the 33 families, which is why the pairwise cosine similarities reported in Subsection III-F take negative as well as positive values. Several descriptors are conceptually related (shape compactness and whitespace ratio; character entropy and visual complexity), and their intercorrelation was not re-estimated in the present study. The eleven tests in Subsection III-C are therefore not independent, which makes the Bonferroni correction conservative and is one reason the Benjamini–Hochberg procedure is also reported.

TABLE I
MORPHOLOGICAL DESCRIPTORS OF THE PARENT STUDY

Descriptor	Property Quantified	Typographic Relevance
Shape compactness	Proportion of the bounding area occupied by the glyph	Sensitive to serifs and to width of construction
Stroke density	Amount of ink per unit of glyph area	Weight and colour of the text block
Stroke contrast	Variation between thickest and thinnest strokes	Modulation; separates old-style, transitional and modern serifs from monoline sans
Character openness	Amount of counter space open to the exterior	Aperture magnitude (not direction)
Whitespace ratio	Proportion of non-ink area within the glyph frame	Spacing and internal space
Character symmetry	Degree of mirror symmetry about the vertical axis	Geometric versus calligraphic construction
Aspect ratio	Height-to-width ratio of the glyph bounding box	Condensed versus extended proportions
Curvature	Proportion and sharpness of curved contour segments	Organic versus geometric construction
Stroke direction	Dominant orientation of stroke edges	Stress axis and construction (partial)
Character entropy	Information entropy of the glyph image	Irregularity of form
Visual complexity	Density of contour detail relative to glyph size	Ornament and detail; display versus text faces

D. Script-Category Coding and Analytical Unit

Each family was coded as serif or sans-serif on the basis of terminal treatment, following standard typographic convention; display faces with modulated terminals were coded by the dominant treatment of the lowercase alphabet. All statistical comparisons are conducted at family level ($n = 33$; 16 historical, 17 contemporary), not at glyph level. This is deliberate. The 36 glyphs comprising a family are not independent observations: they share a designer, a set of construction principles and a single rendering. Treating them as 1,188 independent cases would constitute pseudoreplication [25] and would inflate significance substantially. Glyph-level observations are used only to form family means, and no inference is drawn at glyph level, so a

hierarchical (mixed-effects) model is not required for the tests reported here; an analysis that retains within-family variation would require one and is identified as future work in Subsection III-M. The family-level choice reduces statistical power considerably and the resulting confidence intervals are correspondingly wide; that width is reported rather than avoided, being an accurate representation of what a 33-family corpus can support.

E. Diagnostics and Statistical Procedure

Five diagnostics are applied. D1 compares assigned period against the documented date of the rendered artefact. D2 cross-tabulates script category against era group, tested with Fisher's exact test (odds ratio with conditional maximum-likelihood 95% confidence interval), and quantifies correspondence between the recovered clusters and both script category and era using the adjusted Rand index [29]; where the groups differ in composition, the morphological comparison is re-run on the script-homogeneous subset. The clusters are those of the parent study, obtained by Ward-linkage hierarchical clustering of the standardised family-level descriptors with Euclidean distance and cut at five clusters, the solution reported in the parent study; the diagnostics here concern how that reported partition relates to script category and era, not the choice of the number of clusters. Robustness of D2 to corpus configuration was assessed without re-clustering, by recomputing Fisher's test and both indices on the reported partition when excluding families with ambiguous or discrepant provenance, excluding the born-digital screen faces, restricting to sans-serif families, and balancing era or script by exhaustive or random subsampling (seed 42). D3 re-evaluates the eleven reported descriptor differences (one test per descriptor, eleven tests in the family) under both Bonferroni family-wise correction and the Benjamini–Hochberg false discovery rate procedure [17] at $\alpha = 0.05$. D4 reports classification performance with its 95% Wilson confidence interval against the majority-class baseline, together with balanced accuracy and macro-F1 derived from the confusion matrix (bootstrap 95% intervals, 10,000 resamples, seed 42) and the distribution of fold-level performance across repeated cross-validation. Classification uses the random forest of the parent study with its configuration unchanged; because the purpose here is to evaluate the stability of the reported estimate rather than to optimise it, no hyperparameter tuning was performed. The headline estimate covers predictions for all 33 families, as the confusion matrix reported in Subsection III-E shows, and stability was assessed with stratified five-fold cross-validation (six to seven families per test fold, era ratio preserved in each fold) repeated over multiple random partitions of the corpus. D5 compares documented derivation with morphological proximity: pairs within the corpus whose relation is documented in foundry or designer sources are compared, in cosine similarity, with the most similar pairs for which no relation is documented. Because the parent study

reports the similarity matrix graphically, values were recovered by mapping each cell's colour to the colour scale of the figure; the recovery reproduces the unit diagonal and the symmetry of the matrix, and its resolution of approximately ± 0.03 is sufficient for the ordinal comparison made here.

Group comparisons in the controlled re-analysis use the Mann–Whitney U test with rank-biserial correlation as effect size, appropriate to the sample size and to non-normal distributions. Family-level descriptor values used in that re-analysis were taken from the per-family rankings of the parent study, which report three decimal places. For shape compactness the family-level recomputation for the full corpus gives $p = 0.014$ ($U = 67.5$, $r = 0.50$), against an originally reported value of 0.001; the difference is larger than rounding would produce and cannot be resolved from the values available. Inference in the present study is therefore based on family-level values wherever they are available, and both values are given where relevant.

III. RESULT AND DISCUSSION

A. Provenance Composition of the Corpus

Table I reports the provenance layer for the sixteen families designated historical. Eleven carry design dates after 1899, using the dates recorded in the corpus itself. Five fall within the declared range but those five are dated by their historical models rather than by the artefacts analysed: Libre Caslon Text was released in 2019 and Libre Baskerville in 2012, and the GFS faces are digital-era re-cuttings. Under file-level provenance, therefore, no family in the historical group is an artefact of the declared period. Every rendered file postdates it, most by centuries.

The eleven post-1899 cases are not of one kind, and Table I records each separately. Five are digitisations of pre-digital designs dated between 1927 and 1957 (Futura, Gill Sans, Times New Roman, Helvetica, Univers); for these the mismatch arises from the definition of historical, which the corpus applied to twentieth-century metal types outside its own declared period. Two are original designs of the early digital era for print (Arial, 1982; Avenir, 1988). Four are born-digital screen faces (Georgia, Verdana, Cambria, Calibri). None of the eleven arises from a mislabelled date alone: even where a corpus date is wrong, the correct date also falls after 1899, so the mismatch lies in the definition of the group rather than in the metadata. In the analyses that follow the corpus labels are retained, since the object of study is the inference the corpus supports as constructed; the sensitivity of the main findings to excluding the born-digital faces and the families with discrepant provenance is reported in Subsection III-B.

TABEL II
PROVENANCE LAYER FOR THE SIXTEEN FAMILIES
DESIGNATED HISTORICAL

Family	Designer / Foundry	Year	File	Status	Script
Garamond	after C. Garamond; cut not recorded	c. 1540	n.r.	undocumented	serif
Libre Caslon Text	P. Impallari (Impallari Type), after W. Caslon (c. 1722)	1610†	2019	revival	serif
Libre Baskerville	P. Impallari, after J. Baskerville (1757)	1757	2012	revival	serif
GFS Didot	Greek Font Society, after F. Didot	1810	n.r.	revival	serif
GFS Bodoni	Greek Font Society, after G. B. Bodoni	1798	n.r.	revival	serif
Times New Roman	S. Morison & V. Lardent, Monotype	1932	n.r.	digitisation	serif
Georgia	M. Carter, Microsoft	1993	n.r.	original (screen)	serif
Cambria	J. Bosma, S. Matteson, R. Nicholas	1996†	2004 / 2007	original (screen)	serif
Futura	P. Renner, Bauer	1927	n.r.	digitisation	sans
Gill Sans	E. Gill, Monotype	1928	n.r.	digitisation	sans
Helvetica	M. Miedinger & E. Hoffmann, Haas	1957	n.r.	digitisation	sans
Univers	A. Frutiger, Deberny & Peignot	1957	n.r.	digitisation	sans
Arial	R. Nicholas & P. Saunders, Monotype	1982	n.r.	original	sans
Avenir	A. Frutiger, Linotype	1988	n.r.	original	sans
Verdana	M. Carter, Microsoft	1996	n.r.	original (screen)	sans
Calibri	L. de Groot, Microsoft	2004	2004 / 2007	original (screen)	sans

Two distinctions follow. Four members of the historical group Georgia, Verdana, Cambria and Calibri are born-digital typefaces designed for screen rendering under hinting constraints that did not exist before the 1990s; they share no production technology with the metal types they are grouped alongside. And two coding discrepancies were identified against published sources: Cambria is recorded in the corpus as 1996, while its documented release is 2004 (1996 corresponds to Verdana); and the name Garamond does not resolve to a single artefact, since digital families bearing the name descend from distinct sources and the corpus does not record which cut was rendered. The historical/contemporary

division indexes stylistic tradition rather than artefact chronology. Correspondence between assigned period and documented release is absent for the entire group.

B. Compositional Comparability of the Era Groups

Script-category coding returns a marked asymmetry between the two groups.

TABEL III
SCRIPT-CATEGORY COMPOSITION OF THE TWO ERA GROUPS

Group	Serif	Sans-serif	% Serif
Historical (n = 16)	8	8	50.0
Contemporary (n = 17)	1	16	5.9

Fisher's exact test returns $p = 0.0066$, with an odds ratio of 14.7 (95% CI 1.5–753). The two groups differ significantly in script-category composition before any morphological feature is measured. The interval is wide, as expected with a single contemporary serif family, but its lower bound excludes parity.

The cluster structure recovered from the morphological features tracks this composition rather than period (Fig. 2), and the imbalance is apparent on inspection of the rendered corpus itself (Fig. 3). Cluster 4 (n = 9) contains all nine serif families in the corpus: the eight historical serif faces and Mash, a serif display face assigned to the contemporary group. Cluster 5 (n = 17) contains seven of the eight sans-serif families from the historical group alongside ten contemporary families. Cluster 3 (n = 5) groups Futura with four contemporary geometric and condensed sans faces (Buman, Maneuver, Telemachus and ZT Ravigsfen); Clusters 1 and 2 are the singletons Roadside Sans and Thernaly. No cluster contains both serif and sans-serif families.

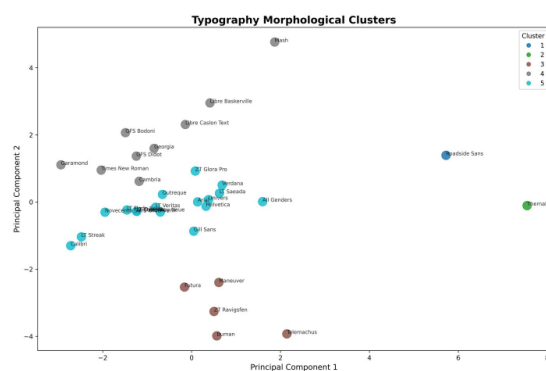


Fig. 2. Morphological cluster assignment projected onto the first two principal components. Cluster 4 (upper left) is the serif cluster; Cluster 5 is the sans-serif cluster, containing families from both era groups.

Quantified against the two candidate groupings, the correspondence is decisive: adjusted Rand index against script category = 0.534, against era = 0.090. The partition is approximately six times better explained by the serif/sans distinction than by period. The same structure appears in the

TABLE V
REPORTED SIGNIFICANCE VALUES UNDER MULTIPLE-COMPARISON
CORRECTION

Descriptor	Raw p	Bonferroni	BH-FDR	Survives
Shape compactness	0.001	0.011	0.011	both
Curvature	0.004	0.044	0.022	both
Visual complexity	0.013	0.143	0.048	FDR only
Stroke density	0.021	0.231	0.053	—
Whitespace ratio	0.028	0.308	0.053	—
Character entropy	0.029	0.319	0.053	—
Character openness	0.037	0.407	0.058	—
Aspect ratio	0.067	0.737	0.092	—
Stroke contrast	0.081	0.891	0.099	—
Character symmetry	0.146	1.000	0.161	—
Stroke direction	0.213	1.000	0.213	—

Note: Bonferroni threshold = $0.05/11 = 0.00455$; BH-FDR = Benjamini–Hochberg false discovery rate.

Under Benjamini–Hochberg false discovery rate control, three survive. Under family-wise Bonferroni correction, two survive: shape compactness and curvature. These two are not independent of the compositional finding above. Shape compactness measures the proportion of the bounding area occupied by the glyph, a quantity directly sensitive to the presence of serifs; curvature distinguishes organic from geometric construction, the principal axis separating serif text faces from geometric sans. The differences that survive correction are precisely those most confounded with script category. Two qualifications apply. First, the values in Table V are those originally reported; the family-level recomputation available for shape compactness gives $p = 0.014$, which would not survive Bonferroni correction (adjusted $p = 0.154$). Second, the parent study reports p -values only, so effect sizes for the eleven descriptors cannot be recovered; for the descriptors with available family-level values, rank-biserial effect sizes are reported in Table VI.

D. Era Differences Controlling for Script Category

The controlling test restricts the comparison to sans-serif families only 8 historical against 16 contemporary holding script category constant. Family-level values were taken from the published per-family rankings (Fig. 6).

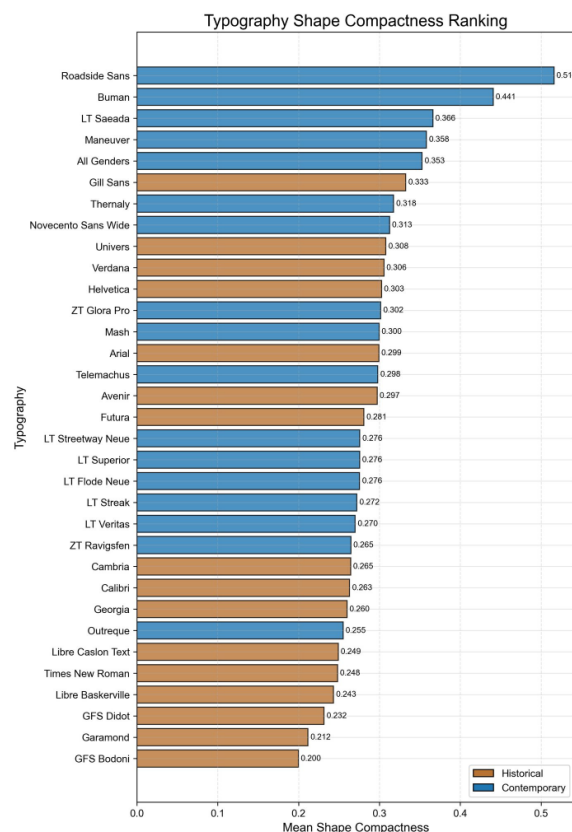


Fig. 6. Per-family shape compactness ranking, coloured by assigned era.

TABLE VI
ERA COMPARISON FOR THREE DESCRIPTORS, FULL CORPUS VERSUS SANS-SERIF FAMILIES ONLY

Descriptor	Full Corpus p	Sans Only p	r	Hist. Median	Cont. Median
Shape compactness	0.014	0.878	0.047	0.301	0.300
Visual complexity	0.094	0.976	0.016	0.497	0.485
Stroke contrast	0.084	0.713	0.102	0.522	0.516

Note: Mann–Whitney U, two-sided; r = rank-biserial correlation.

The era differences vanish across all three descriptors for which per-family values are published. Shape compactness the single descriptor that survived family-wise correction in the full corpus returns $p = 0.878$ with a median difference of 0.001 once serif families are removed. Visual complexity and stroke contrast behave identically. Effect sizes fall below 0.11 in all three cases. The test covers three of the eleven descriptors those for which family-level values are reported in the published rankings. Curvature, the second descriptor to survive family-wise correction, is not among them; extension of the controlled comparison to the full descriptor set is required before the result can be generalised beyond the three tested.

The era groups differ significantly in script-category composition, and that composition accounts for the observed morphological differences. Within a script-homogeneous

subset, the historical and contemporary groups are morphologically indistinguishable on the descriptors tested.

E. Separability and Stability

Classification of era from the eleven descriptors was performed at family level with a random forest (Fig. 7). The confusion matrix sums to 33 ($13 + 3 + 5 + 12$), confirming the analytical unit. Accuracy was $25/33 = 75.8\%$, with a 95% Wilson confidence interval of $[59.0\%, 87.2\%]$, against a majority-class baseline of 51.5%: the accuracy obtained by assigning every family to the larger, contemporary class (17/33). Because the two classes are nearly balanced, accuracy is not materially inflated by imbalance, but class-sensitive metrics and interval estimates are reported alongside it, as recommended for small-sample evaluation [35]: balanced accuracy is 0.759 (bootstrap 95% CI 0.61–0.90) and macro-F1 0.757 (0.60–0.88), with recall of 0.81 for historical and 0.71 for contemporary families and precision of 0.72 and 0.80 respectively. A one-sided exact binomial test against the baseline gives $p = 0.004$, and the lower bound of the accuracy interval lies 7.5 percentage points above baseline; the point estimate is compatible with a range extending from marginal improvement over chance to substantial separability, and the data do not discriminate between these.

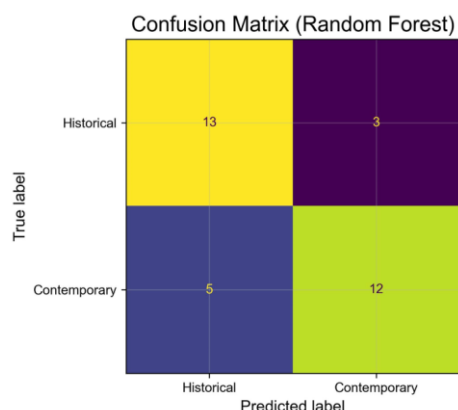


Fig. 7. Confusion matrix for the best-performing classifier.

Mean fold-level F1 is approximately 0.64 and mean ROC-AUC approximately 0.74. With six or seven families per test fold, a single misclassification shifts fold accuracy by 14–17 percentage points, and the minimum of 0.29 corresponds to two correct predictions out of seven. This dispersion is therefore the expected consequence of sample size rather than an anomaly; the gap between the headline estimate and the repeated-validation mean is the pattern that small-sample validation studies predict when a single favourable partition is reported [26].

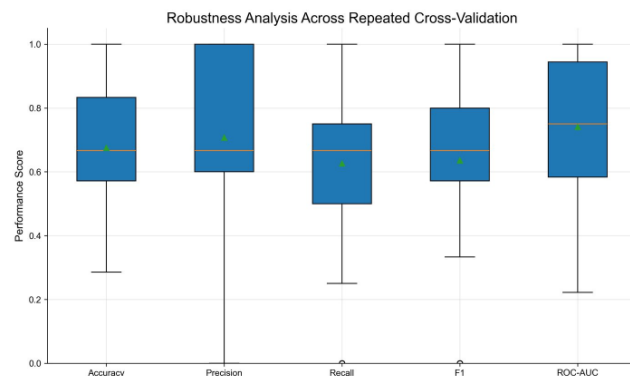


Fig. 8. Distribution of performance metrics across repeated cross-validation. The spread on every metric extends far below the reported headline values.

Two further observations corroborate instability rather than performance. Ensemble methods underperform the single model: soft voting returns 0.667 and stacking 0.545 against the base model's 0.756. Ensembling ordinarily improves or matches base performance; a decline of this magnitude indicates the base estimate is not robust. Separately, one classifier returns approximately 70% accuracy alongside an ROC-AUC of 0.49 indistinguishable from random ranking. Morphological separability between the era groups is weak and unstable. The headline estimate is a single favourable draw from a distribution whose central tendency lies close to baseline.

F. Recurrent Attractors in the Inferred Genealogy

Assignment of each contemporary family to its most similar historical family produces a markedly uneven distribution (Fig. 9). Three families account for twelve of the seventeen links: Calibri (5), Univers (4) and Futura (3). The remaining five are distributed singly across Arial, Avenir, Gill Sans, Libre Baskerville and Verdana.

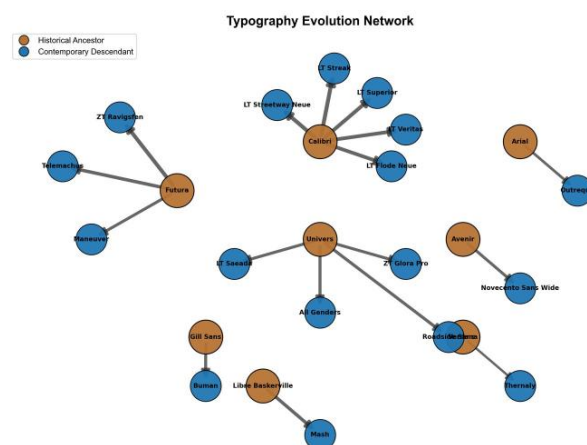


Fig. 9. Inferred evolution network. Calibri, Univers and Futura account for twelve of the seventeen assigned ancestor–descendant links.

TABLE VII
INFERRED ANCESTOR ASSIGNMENTS WITH SCRIPT CATEGORY OF
ANCESTOR AND DESCENDANTS

Inferred Ancestor	Release	Script	Descendants	Descendant Script
Calibri	2004	sans	5	all sans
Univers	1957	sans	4	all sans
Futura	1927	sans	3	all sans
Arial	1982	sans	1	sans
Avenir	1988	sans	1	sans
Gill Sans	1928	sans	1	sans
Verdana	1996	sans	1	sans
Libre Baskerville	2012	serif	1	serif

The pattern is exact: every assignment matches ancestor and descendant on script category, with no exceptions across all seventeen links. Sixteen of the seventeen link a sans ancestor to a sans descendant; the single serif ancestor, Libre Baskerville, is assigned the corpus's single contemporary serif face.

Provenance sharpens the point. Calibri was released in 2004 18 years before the earliest family in the contemporary group and 464 years after the oldest family in its own. Positioned by documented date rather than by assigned label, it sits far closer to the group it is said to have generated than to the group it is said to belong to. Its recurrence as inferred ancestor reflects a central position in the morphological feature space occupied by contemporary sans-serif design, a position it holds because it is itself a recent humanist sans.

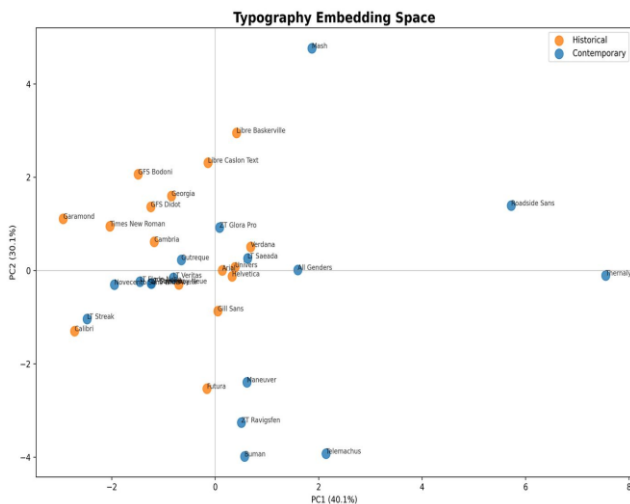


Fig. 10. Labelled embedding of the 33 families. Calibri lies at the periphery of the sans-serif region alongside the contemporary LT family, far from the serif text faces with which it is grouped by era label.

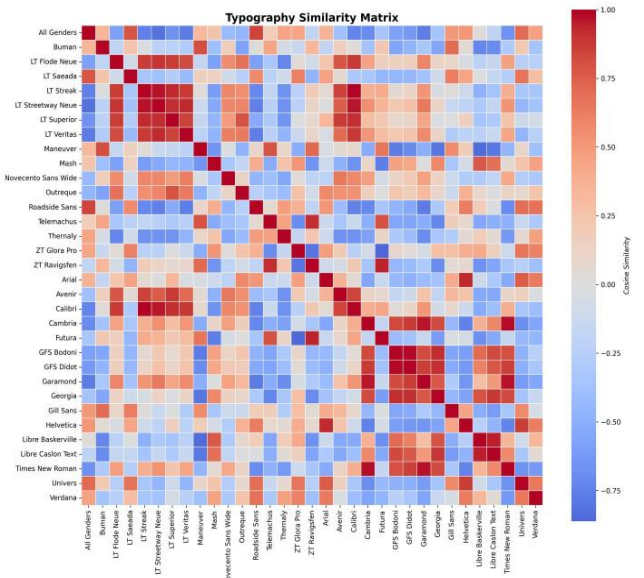


Fig. 11. Pairwise morphological similarity across the 33 families. The dominant block structure corresponds to serif and sans-serif groupings rather than to era.

G. Documented Derivation and Morphological Proximity

Table VIII comparing pairs whose relation is documented with pairs whose relation is not. Two documented derivations are present in the corpus. Arial was drawn by Monotype with character widths identical to those of Helvetica, so that it could substitute for it without altering line breaks; whatever its debt to Monotype's own grotesques, its metrics are a documented derivation from Helvetica. Avenir was designed by Adrian Frutiger explicitly in the tradition of Erbar and Futura, as a humanised reinterpretation of the geometric sans [27]. A third pair, Helvetica and Univers, released in the same year by different foundries, shares a background in the grotesque tradition but involves no transmission between the two designs [27].

TABLE VIII
DOCUMENTED AND UNDOCUMENTED PAIRS IN MORPHOLOGICAL
SIMILARITY

Pair	Documented Relation	Cosine Similarity	Percentile
Helvetica – Arial	derivation (metrics)	0.92	97.7
Futura – Avenir	reinterpretation (designer)	0.08	57.8
Helvetica – Univers	parallel, no transmission	0.87	95.1
Calibri – LT Streak	none	0.99	99.8
Cambria – Times New Roman	none	0.97	99.4
Calibri – LT Streetway Neue	none	0.96	98.7
Futura – ZT Ravigsfen	none	0.94	98.3
Libre Baskerville – Mash	none (inferred link)	0.77	90.5

The comparison answers RQ4 in the negative. The documented derivation of Arial from Helvetica does produce high similarity (0.92, 97.7th percentile), but four pairs with no documented relation reach higher values, including Calibri and LT Streak at 0.99, the highest off-diagonal similarity in the corpus. Conversely, Avenir, which its designer placed explicitly in the tradition of Futura, sits at 0.08, close to the corpus median and only twelfth among Futura's neighbours, while Futura's nearest neighbour is ZT Ravigsfen, a contemporary display face with no documented relation to it. Similarity is therefore neither sufficient for documented descent, since high-similarity pairs occur without any relation, nor necessary for it, since a documented reinterpretation lies at median similarity. Documented pairs within a 33-family corpus are few, and the comparison is illustrative rather than inferential; its force lies in the existence of counter-examples in both directions, which a similarity-based genealogy cannot accommodate.

H. Close Reading of the Divergent Cases

The statistical results identify where the morphological account and the documentary record diverge. This subsection reads those divergences as a designer would, examining the letterforms themselves (Figs. 1 and 2) to establish what the measurements registered and what they passed over.

1) *The Calibri Lineage*: Calibri and its five assigned descendants share a recognisable formal register: open apertures, moderate stroke contrast, humanist proportions, and terminals that resolve softly rather than at a sharp cut. To a reader of specimens the resemblance is real, not an artefact. But the resemblance has two distinct causes. Calibri's softened corners were drawn as a technical accommodation to sub-pixel rendering, a constraint specific to early-2000s screen display. The contemporary faces arrive at comparable softness through a different route: a screen-first display idiom in which optical smoothness signals contemporaneity. The forms converge; the reasons do not. Reading the group as a lineage attributes to transmission what is better explained as two independent responses to the condition of being rendered on a screen.

2) *Mash and the Limits of a Single Ancestor*: Mash is the corpus's only contemporary serif, and the only contemporary family assigned to the serif cluster. Its assigned ancestor is Libre Baskerville. Visually, however, Mash reads as a Didone-adjacent display face: pronounced thick-to-thin modulation, fine unbracketed serifs, vertical stress. A type historian comparing specimens would place it nearer Bodoni or Didot than Baskerville, whose transitional serifs are bracketed and whose contrast is markedly gentler. The assignment is not arbitrary Mash and Libre Baskerville do share compactness and stroke density but the descriptors that produced the match are insensitive to the features on which the historical judgement turns.

3) *Novecento Sans Wide, a Revival inside the Contemporary Group*: This family registers the lowest visual complexity in the entire corpus (0.320), below every historical face. Its formal language wide geometric capitals, minimal modulation, monoline construction belongs to early twentieth-century lettering rather than to contemporary display practice. Grouped with the contemporary families by label, it functions in the feature space as an outlier pulling toward the geometric sans of the 1920s. It is the mirror of the screen faces sitting in the historical group: a revival misfiled as new, where those are new designs misfiled as old.

4) *The Singletons*: Roadside Sans and Thernaly each occupy a cluster alone. Both are heavy display faces with extreme proportions Roadside Sans broad and rounded, Thernaly narrow and compressed. Their isolation is usually read as noise. It is better read as a property of the object: display typography is designed to be distinctive, and distinctiveness is precisely what resists placement in a shared feature space. A taxonomy built on aggregate morphology will always place the most deliberately singular designs at its margins, which is a statement about the taxonomy rather than about the designs.

I. Convergence Is Not Descent

The findings converge on a single methodological point: morphological similarity measures formal affinity, and formal affinity is a necessary but not sufficient condition for historical descent. Two typefaces may arrive at similar forms without any transmission between them. Screen rendering constraints in the 1990s pushed independent designers toward open apertures, large x-heights and simplified terminals; industrial legibility standards produced convergent solutions across unrelated foundries; geometric construction yields similar skeletons whenever it is adopted, irrespective of lineage. In each case similarity records a shared response to a shared constraint, and reading it as ancestry inverts the causal direction. The documented-pair comparison (Table VIII) shows both failure modes within this corpus. The result parallels the central difficulty of phylogenetic analysis of material culture, where resemblance produced by convergence or borrowing cannot be separated from inherited resemblance on the basis of the traits alone [23], [24]. It also clarifies why learned font similarity serves classification, retrieval and generation well [20]–[22], [31], [32]: in those applications proximity is appropriate precisely because no historical claim is made.

Type history supplies the demonstration. The types issued as Garamond resembled sixteenth-century French sources closely enough to sustain the attribution for three centuries; the correction came from matrices and archival records rather than from the letterforms, which had been continuously available for inspection [4]. Formal analysis had every opportunity and did not detect the error. A computational genealogy operating on form alone occupies the same epistemic position, with greater speed and larger corpora.

J. Corpus Composition as a Hidden Variable

The compositional asymmetry is the study's most consequential finding, because it is not confined to this corpus. Where a corpus is partitioned for comparison, any attribute differing systematically between partitions and correlated with the measured features will be absorbed into the reported difference. Here the partition is era; the differing attribute is script category; the correlation with morphological descriptors is direct, since serif presence is among the most powerful determinants of compactness, curvature and stroke contrast. The controlled re-analysis confirms this directly: removing serif families removes the effect.

The historical group here is 50% serif because historical text typography is substantially serif; the contemporary group is 94% sans-serif because contemporary display typography is substantially sans-serif. The imbalance is not a sampling error it reflects the field. But that is precisely why it must be reported: a genuine difference between eras in what typography is for becomes, when passed through a morphological pipeline, a spurious difference in what typography looks like as a function of period.

K. Centrality Is Not Influence

A family assigned as ancestor to many descendants may occupy a central position in feature space rather than a generative position in history. Under maximum-similarity assignment, the family nearest the centroid of a dense region attracts disproportionate links, since it minimises distance to many candidates simultaneously. Centrality is a property of the feature distribution; influence is a property of the historical record. The two coincide only when the corpus is constructed so that historically influential designs also occupy central positions an assumption requiring justification, not stipulation. That all seventeen assignments match on script category, with none crossing the serif/sans boundary, indicates the assignment is resolving script category first and any residual signal second.

L. What the Descriptor Set Does Not Measure

The divergences above share a structure. In each case the descriptors registered quantity properties how much ink, how compact the form, how symmetric, how varied while the historical judgement turned on quality properties: what kind of serif, what axis of stress, what shape of terminal.

Four features central to typographic classification are absent from the eleven descriptors. Serif form is not among them: the distinction between bracketed serifs, which join the stem on a curve, and unbracketed hairline serifs, which meet it abruptly, separates transitional from modern faces and is the single most consequential distinction in serif type history. Stress axis is not measured: whether the thin points of a curve fall on an inclined or a vertical axis is the primary organising principle of the Vox-ATypeI serif categories. Terminal shape is not captured: ball, teardrop, sheared and flat terminals are diagnostic of period and of foundry. And aperture is measured only as magnitude, not as direction openness registers how

much counter space exists, not whether the counters open outward or close inward, which is what distinguishes a humanist from a grotesque sans.

This is not a deficiency in the descriptors as designed. Compactness, entropy and symmetry are well-formed measures that generalise across scripts and require no typographic expertise to compute which is exactly why they are useful for a machine. But their generality is purchased by abstracting away the features that carry historical information. Type history is written in the details of the joint, the terminal and the axis; a descriptor set that measures the glyph as a distribution of ink cannot reach them. The morphological genealogy diverges from the documented genealogy not because the computation is faulty but because the two are reading different properties of the same object.

M. The 'a' and the 'g'

A related loss occurs through aggregation. Family-level values are means across 36 glyphs, which treats every character as an equal contributor to the family's signature. Typographic identification does not work this way. Specimen books and type-identification manuals lead with the lowercase 'a' and 'g' because those two letters carry disproportionate diagnostic weight: whether the 'g' is single- or double-storey, whether the 'a' has a straight or curved spine, whether the bowl is small and the aperture tight or the reverse.

Fig. 1 makes the point directly. Set side by side, the 'a' and 'g' of Garamond, Bodoni and Futura are immediately separable by anyone with basic typographic training small bowl and inclined stress in the first, high contrast and ball terminal in the second, single-storey geometric construction in the third. Under family-level aggregation each of these letters contributes one thirty-sixth of the signal, and the numerals ten of the thirty-six glyphs, and among the least stylistically diagnostic characters in a Latin font contribute more than a quarter.

This suggests a concrete direction for extension rather than a limitation to be conceded. A weighted or per-glyph analysis, giving the diagnostic letters the standing that typographic practice already gives them, would test whether lineage information survives at the level where designers actually locate it. The present corpus contains the material for that test; only the aggregation discards it.

N. Function Rather than Period

The compositional asymmetry documented in Subsection III-B admits an interpretation beyond the statistical one. The historical group is half serif because the typography of that period, as represented in any standard corpus, is dominated by text faces designed for continuous reading in print. The contemporary group is almost entirely sans-serif because the families available for sampling from contemporary independent foundries are overwhelmingly display faces designed for screen, brand and short-form use.

The two groups therefore differ in purpose before they differ in period. Text typography is constrained toward

legibility at small sizes, moderate contrast and unobtrusive form; display typography is rewarded for the opposite. What a morphological pipeline records as a difference between eras is, read through the artefacts themselves, a difference between functions that the sampling has aligned with periods.

This reframes the parent framework's finding rather than contradicting it. Continuity and adaptation are real features of typographic history, and the visual heritage the framework identifies is genuine. But the axis along which the corpus varies most strongly is function, not chronology and a study that wishes to speak about historical evolution must either sample text and display typography from both periods, or state that its claims are conditional on the functional composition it happens to contain.

O. What the Framework Can Legitimately Claim

None of this indicates that morphological analysis of typography is unproductive. The descriptors are well-specified, the rendering pipeline is controlled, and the resulting space demonstrably organises typefaces in ways corresponding to recognised typographic categories the recovery of a serif/sans partition at ARI 0.534 is itself a validation of the feature set. What requires revision is the evidentiary vocabulary attached to the output. A similarity matrix supports statements about formal proximity. A cluster structure supports statements about the organisation of a feature space. Neither supports statements about transmission, inheritance or descent without independent evidence.

P. A Reporting Protocol for Computational Typography

Five reporting requirements follow from these findings for any computational study that offers morphological similarity as evidence about typographic history:

- Report the artefact, not the name font file, version and foundry, since a typeface name does not identify a unique design.
- Report documented release dates, distinguishing the digital artefact from its historical model.
- Report corpus composition along any dimension correlated with the measured features minimally script category for Latin corpora.
- Report assignment strength, not only the top match, and treat low-margin assignments as unresolved.
- State the evidentiary claim explicitly whether similarity is offered as evidence of affinity or of descent.

The protocol imposes documentation cost and no analytical cost. It does not constrain what may be measured, only what may be claimed on the basis of measurement. It applies directly to other computational studies of Latin-script typography. For non-Latin scripts the composition dimension in the third requirement must be redefined, since the serif/sans distinction has no direct counterpart; for other visual corpora the second requirement generalises to distinguishing the analysed reproduction from the historical object it stands for. The protocol does not by itself establish or refute descent; it

specifies the conditions under which a similarity-based claim can be evaluated.

Q. Limitations

The corpus comprises 33 Latin-script families in a single weight, and the findings do not extend to non-Latin writing systems, where the serif/sans distinction has no counterpart. The controlled re-analysis in Subsection III-D covers three of the eleven descriptors those for which per-family values are published and extension to the remaining eight would strengthen the conclusion, though the three tested include shape compactness, one of the two descriptors that survived family-wise correction; curvature, the other survivor, was not among the descriptors for which per-family values are published and remains untested. As set out in Subsection III-L, descriptors are aggregated across 36 glyphs per family, discarding the per-glyph variation on which typographic identification conventionally relies. Provenance documentation is uneven: canonical text faces are extensively documented while independent-foundry display releases frequently are not, and this asymmetry mirrors the compositional imbalance the study identifies. The analysis is also conditional on a classification system its own governing body has withdrawn.

IV. CONCLUSION

This study tested how far morphological similarity can support claims of genealogical descent in a computational typography corpus. On RQ1, the historical/contemporary division does not correspond to artefact chronology: eleven of sixteen historical families carry design dates after the declared period, and no rendered file belongs to it. On RQ2, the two groups differ in script-category composition before any morphology is measured, and that composition, rather than period, organises the recovered structure across every corpus configuration examined. On RQ3, separability between the eras is weak and unstable: two of eleven descriptor differences survive family-wise correction, no era difference is detected within sans-serif families on the descriptors testable, and a headline accuracy of 75.8% falls to about 0.68 under repeated validation against a 51.5% baseline. On RQ4, similarity neither identifies documented derivation nor excludes pairs without any documented relation.

Morphological similarity is therefore evidence of formal affinity, conditional on corpus composition, and is necessary but not sufficient for a claim of descent. The contribution of the study is methodological: an auditable provenance layer, five family-level diagnostics and a reporting protocol that together make such claims evaluable. Future work should recompute the analysis with descriptor-level and algorithm-level robustness checks, extend it to per-glyph and weighted analyses that give diagnostic letters their typographic weight, sample text and display faces from each period, and test the protocol on non-Latin corpora.

REFERENCES

- [1] T. Arnold and L. Tilton, "Distant viewing: Analyzing large visual corpora," *Digital Scholarship in the Humanities*, 2019, doi: 10.1093/llc/fqz013.
- [2] L. Tilton and J. Wigard, "Seeing formalism or formal viewing: Computational formalism for the analysis of visual media forms and contexts," *Journal of Cultural Analytics*, 2025, doi: 10.22148/001c.144827.
- [3] M. Mason, S. Kirchner, and C. Powell, "Automated glyph feature detection using convolutional neural networks," in *Lecture Notes in Computer Science*. Cham, Switzerland: Springer, 2026, doi: 10.1007/978-3-032-06744-9_3.
- [4] P. Beaujon [B. Warde], "The 'Garamond' types: Sixteenth and seventeenth century sources considered," *The Fleuron*, no. 5, pp. 131–179, 1926.
- [5] P. Hiskiawan, C. Chih, C. Zheng, and K. Ye, "Processing of electrical resistivity tomography data using convolutional neural network in ERT-NET architectures," *Arabian Journal of Geosciences*, 2023, doi: 10.1007/s12517-023-11690-w.
- [6] Y. Erlyana and Y. Hansen, "Tinjauan tipografi pada poster film horor Indonesia garapan Rizal Mantovani," *Jurnal RupaRupa*, vol. 3, no. 1, pp. 17–28, 2014.
- [7] Y. Erlyana, Winnie, and Veronica, "Desain layout buku tipografi untuk kemasan dengan metode design thinking," *Jurnal Bahasa Rupa*.
- [8] S. Everlin and Y. Erlyana, "Analisis perubahan desain logo Gojek tahun 2019," *DESKOMVIS*, vol. 1, no. 1, pp. 72–88, 2020, doi: 10.38010/dkv.v1i1.11.
- [9] S. Everlin, Y. Erlyana, and C. Gunawan, "Analisis semiotika brand value melalui artist-brand collaboration Uniqlo UT," *Jurnal Desain*, vol. 9, no. 1, p. 64, 2021, doi: 10.30998/jd.v9i1.10293.
- [10] Y. Erlyana, S. Everlin, and I. F. Yuwono, "Analisis color palette berdasarkan rasa warna sebagai penguat daya tarik emosional dalam video anak," *Andharupa*, vol. 9, no. 3, pp. 396–411, 2023.
- [11] H. Thamrin, "Pengaruh estetika visual dan personal branding terhadap social currency content creator TikTok di Indonesia," *Filosofi*, vol. 1, no. 3, pp. 267–282, 2024, doi: 10.62383/filosofi.v1i3.246.
- [12] H. Thamrin, "Analisa brand identity yang dimediasi oleh packaging design terhadap purchase intention," *Pomdasi Journal of Applied Science and Engineering*, 2024.
- [13] R. Bringhurst, *The Elements of Typographic Style*. Vancouver, Canada: Hartley & Marks,
- [14] P. B. Meggs, *A History of Graphic Design*. New York, NY, USA: John Wiley & Sons
- [15] Association Typographique Internationale, "ATypI de-adopts the Vox-ATypI typeface classification system," 2021. [Online].
- [16] J. Drucker, "Humanities approaches to graphical display," *Digital Humanities Quarterly*, vol. 5, no. 1, 2011.
- [17] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B*, vol. 57, no. 1, pp. 289–300, 1995.
- [18] P. Hiskiawan, E. Stephanie, H. Heryanto, and S. A. Feri, "Trustworthy data science framework for non-invasive nutritional screening using computer vision," in *Proc. ICIMCIS 2025*, 2025, pp. 1744–1748.
- [19] P. Hiskiawan, J. William, and L. F. Tio Jansel, "A hybrid data science framework for forecasting Bitcoin prices," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2089–2101, 2025, doi: 10.30871/jaic.v9i5.10631.
- [20] N. D. F. Campbell and J. Kautz, "Learning a manifold of fonts," *ACM Transactions on Graphics*, vol. 33, no. 4, art. 91, 2014, doi: 10.1145/2601097.2601212.
- [21] E. Balashova, A. H. Bermanno, V. G. Kim, S. DiVerdi, A. Hertzmann, and T. Funkhouser, "Learning a stroke-based representation for fonts," *Computer Graphics Forum*, vol. 38, pp. 429–442, 2019, doi: 10.1111/cg.13540.
- [22] P. O'Donovan, J. Libeks, A. Agarwala, and A. Hertzmann, "Exploratory font selection using crowdsourced attributes," *ACM Transactions on Graphics*, vol. 33, no. 4, art. 92, 2014, doi: 10.1145/2601097.2601110.
- [23] I. Temkin and N. Eldredge, "Phylogenetics and material cultural evolution," *Current Anthropology*, vol. 48, no. 1, pp. 146–153, 2007.
- [24] J. Tehrani and M. Collard, "Investigating cultural evolution through biological phylogenetic analyses of Turkmen textiles," *Journal of Anthropological Archaeology*, vol. 21, no. 4, pp. 443–463, 2002.
- [25] S. H. Hurlbert, "Pseudoreplication and the design of ecological field experiments," *Ecological Monographs*, vol. 54, no. 2, pp. 187–211, 1984.
- [26] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLoS ONE*, vol. 14, no. 11, e0224365, 2019, doi: 10.1371/journal.pone.0224365.
- [27] H. Osterer and P. Stamm, Eds., *Adrian Frutiger – Typefaces: The Complete Works*. Basel, Switzerland: Birkhäuser, 2009.
- [28] Microsoft Corporation, "Calibri font family" and "Cambria font family," *Microsoft Typography documentation*. [Online]. Available: <https://learn.microsoft.com/typography/font-list/calibri> (accessed Sep. 28, 2026).
- [29] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.
- [30] D. G. Altman and J. M. Bland, "Absence of evidence is not evidence of absence," *BMJ*, vol. 311, no. 7003, p. 485, 1995.
- [31] I. Memon, A. u. H. Muhammad, and J. Choi, "Robustness of contrastive learning on multilingual font style classification using various contrastive loss functions," *Applied Sciences*, vol. 13, no. 6, art. 3635, 2023, doi: 10.3390/app13063635.
- [32] A. u. Hassan, I. Memon, and J. Choi, "Real-time high quality font generation with Conditional Font GAN," *Expert Systems with Applications*, vol. 213, art. 118907, 2023, doi: 10.1016/j.eswa.2022.118907.
- [33] T. Arnold and L. Tilton, *Distant Viewing: Computational Exploration of Digital Images*. Cambridge, MA, USA: MIT Press, 2023.
- [34] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, art. 100804, 2023, doi: 10.1016/j.patter.2023.100804.
- [35] G. Varoquaux and O. Colliot, "Evaluating machine learning models and their diagnostic value," in *Machine Learning for Brain Disorders*, O. Colliot, Ed. New York, NY, USA: Humana/Springer, 2023, doi: 10.1007/978-1-0716-3195-9.