Data and Metadata. 2025; 4:1062
doi: 10.56294/dm20251062

ORIGINAL

Predictive Models of Typographic Preference in Digital Media

Modelos Predictivos de Preferencia Tipográfica en Medios Digitales

Ana Lucía Rivera-Abarca[1], Jazmín Isabel García-Guerra[1], Héctor Oswaldo Aguilar-Cajas[1], Heidy Elizabeth Vergara-Zurita[1], José Israel López-Pumalema[2], Freddy Armijos-Arcos[3]

[1]Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba. Ecuador.
[2]Universidad Estatal Península de Santa Elena.
[3]Ingeniería, Diseño y Consultoría (IDYC Cia. Ltda.)

Cite as: Rivera-Abarca AL, García-Guerra JI, Aguilar-Cajas HO, Vergara-Zurita HE, López-Pumalema JI, Armijos-Arcos F. Predictive Models of Typographic Preference in Digital Media. Data and Metadata. 2025; 4:1062. https://doi.org/10.56294/dm20251062

Submitted: 04-10-2024 **Revised:** 18-02-2025 **Accepted:** 09-06-2025 **Published:** 10-06-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding author: Ana Lucía Rivera-Abarca

ABSTRACT

Introduction: this article explores how typography influences user experience in digital environments, highlighting its evolution from the 11th century to the Internet era.
Objective: the aim of this research was to examine the psychological impact of fonts, which evoke emotional responses and affect readability, design and user behavior.
Method: predictive models, such as regression, classification and time series, are used to analyze typographic preferences, helping designers to optimize digital interfaces.
Results: the study simulated data from 1 000 participants, considering variables such as age, gender, educational level and context of use, revealing a predominant preference for Sans Serif typefaces (63,3 %), especially in academic reading. The Logistic Regression and SVM models showed a moderate performance (accuracy of 0,627 and 0,634), with better ability to identify preferences for Sans Serif, although with limitations for the minority class (Serif).
Conclusion: it was concluded that psychological, cultural and contextual factors significantly influence preferences, highlighting the need to integrate these variables in future models to improve accuracy and personalization in digital design.

Keywords: Typography; Predictive Modeling; Digital Media; User Preferences; Font Psychology.

RESUMEN

Introducción: este artículo explora cómo la tipografía influye en la experiencia del usuario en entornos digitales, destacando su evolución desde el siglo XI hasta la era de Internet.
Objetivo: el objetivo de la presente investigación fue examina el impacto psicológico de las fuentes, que evocan respuestas emocionales y afectan la legibilidad, el diseño y el comportamiento del usuario.
Método: se emplean modelos predictivos, como regresión, clasificación y series temporales, para analizar preferencias tipográficas, ayudando a diseñadores a optimizar interfaces digitales.
Resultados: el estudio simuló datos de 1 000 participantes, considerando variables como edad, género, nivel educativo y contexto de uso, revelando una preferencia predominante por tipografías Sans Serif (63,3 %), especialmente en lectura académica. Los modelos de Regresión Logística y SVM mostraron un rendimiento moderado (precisión de 0,627 y 0,634), con mejor capacidad para identificar preferencias por Sans Serif, aunque con limitaciones para la clase minoritaria (Serif).
Conclusión: se concluyó que los factores psicológicos, culturales y contextuales influyen significativamente

en las preferencias, subrayando la necesidad de integrar estas variables en futuros modelos para mejorar laprecisión y personalización en el diseño digital.

Palabras clave: Tipografía; Modelos Predictivos; Medios Digitales; Preferencias de Usuario; Psicología de Fuentes.

INTRODUCTION

Typography, considered an art form, has a fascinating history dating back to the 11th century when it was mainly used in printed materials such as books and magazines. However, with the advent of the Internet, everything changed dramatically. The digital world opened new doors for typography, taking it beyond paper and allowing broader access to various typefaces.[1,2]

As digital media grew, so did the range of typefaces: classic fonts such as Times New Roman, Arial, and Helvetica emerged, and newer, more casual styles such as Comic Sans. This evolution expanded the choices available and prompted a constant reflection on how typography influences the user's experience in the digital environment.[3,4]

The psychological impact of typography plays a crucial role in how users interact with digital content. The psychology of fonts reveals that different fonts can evoke different emotional responses, influencing perceptions, decision-making, and behavior.[5,6] The selection of typefaces, layout, and size significantly affects the legibility and overall aesthetic appeal of digital designs, making it a critical consideration in web design.

Predictive modeling has emerged as a technique for analyzing user preferences in typography. This statistical method leverages historical data to predict possible outcomes related to user interaction with different typographic elements, helping designers create more effective and engaging digital experiences.[7] By understanding user behavior about typographic choices, designers can improve visual consistency and foster emotional connections with their audience.[8] Therefore, the integration of predictive models in typography refines aesthetic choices and optimizes user interaction and satisfaction in digital media.

METHOD

Statistical models

Predictive modeling

Predictive modeling was selected because it plays a crucial role in understanding typographic preferences in digital media. Using statistical techniques, data patterns can be analyzed to predict user preferences and behaviors regarding typography. This approach allows researchers and designers to predict how different typographic choices influence user engagement and satisfaction.

Types of predictive models

Several types of predictive models can be applied in typography, each with unique purposes. Still, regression models were used because they can predict continuous outcomes, such as a user's overall satisfaction rating based on specific typographic characteristics. By examining the relationship between the independent variables (such as font size, style, and spacing) and the dependent variable (user satisfaction), these models can reveal which typographic elements most significantly affect user perceptions.

Classification models

Classification models were also used to categorize the results into distinct groups. In typography, these models can predict whether a user will prefer one font over another based on their previous interactions and choices. For example, by estimating the likelihood that a user will choose a particular typeface, designers can tailor their offerings to meet user preferences more effectively.
Time series models

Time series models

Time series models analyzing data points collected over time were also evaluated, making them particularly useful for observing trends in typographic preferences. These models can help identify seasonal variations, such as changes in user preferences during certain times of the year, which could inform design strategies during peak periods such as holidays or events.

Digital media applications

The decision to use predictive modeling in typography was made because it can significantly enhance the user experience by allowing designers to create more personalized and engaging digital interfaces. By leveraging the insights gained from these models, organizations can make informed decisions about typography that cater to the preferences of their target audience, ultimately increasing user engagement and satisfaction. Integrating predictive models into the understanding of typographic preferences helps create visually appealing content and aligns with the psychological aspects of design, ensuring that text resonates well with users. This holistic approach is vital to maximizing the effectiveness of digital media in an increasingly competitive landscape.

Factors influencing typographic preference

Several key factors were considered to contribute to typographic preference among users. Psychological impact The psychological influence of typography is profound, shaping brand perception and emotional response. Typography affects the tone of communication and contextual relevance, which in turn influences how users interact with content and the role users often have personal histories and cultural backgrounds that shape their perceptions of typefaces, making typography a deeply subjective experience. This nuanced interaction between type and emotion highlights the importance of considering psychological factors when selecting typography for digital media.

User behavior and decision-making

Factors such as visual hierarchy and emotional connection to content were considered in user behavior and decision-making processes. Cognitive biases and emotional appeals also affect how users respond to different typefaces, as they can evoke specific feelings and associations that influence user behavior.

Emotional design considerations

The concept of emotional design was evaluated as it is crucial when assessing typographic preferences. This approach emphasizes the creation of designs that elicit positive feelings in users, thereby increasing engagement and fostering a sense of connection. Emotional design integrates color, typography, and narrative elements to create memorable interactions. Successful examples, such as Airbnb, demonstrate how effective typographic choices can enhance feelings of trust and belonging, which are central to the user experience.

Cultural and contextual factors

Cultural background was also evaluated, as it significantly influences typographic preference. Users from different regions may respond differently to specific typefaces based on their cultural histories, and people's perceptions of typography are influenced by their cultural contexts, suggesting that designers should consider these variations when making typographic choices.

Data used

For this study, a simulation model was developed to analyze typographic preferences in digital environments, using a simulated sample of 1 000 participants. Data generation was based on carefully selected probability distributions: a normal distribution ($\mu=35$, $\sigma=12$) for age, bounded between 18 and 70 years; categorical distributions for gender (48 % male, 48 % female, 4 % other), educational level (30 % high school, 50 % university, 20 % graduate), and context of use (60 % recreational reading, 40 % academic reading). Time spent was modeled using an exponential distribution with a mean of 10, limited to a maximum of 60 minutes, reflecting realistic patterns of digital content consumption.

The dependent variable, typographic preference (Sans-serif vs. Serif), was generated using a probabilistic model incorporating multiple influencing factors. This model assigns a base probability of 0,5, which is modified according to specific user characteristics: age under 30 increases the likelihood of preference for Sans-serif by 0,1, use of mobile devices increases it by 0,15, and academic background adds 0,1.

Conservatively, factors such as age over 50, postgraduate education level, and extended dwell time (>30 minutes) reduce the probability by 0,1, 0,05, and 0,1, respectively, building on previous findings in the literature on legibility and typographic preferences in digital media.

RESULTS

Comparative analysis of the classification models reveals moderate performance in predicting users' typographic preferences. Logistic Regression and Support Vector Machine (SVM) exhibit slightly higher cross-validation accuracies, with values of 0,627±0,052 and 0,634±0,063, respectively. These results suggest a similar predictive ability between both models in terms of overall accuracy, as can be seen in table 1.

Table 1. Comparative analysis of classification models

Statistician	Age	Permanence	Sans_Serif preference
Sample	1000	1000	1000
Mean	35,08	10,09	0,63
Deviation	11,04	10,11	0,48

https://doi.org/10.56294/dm20251062

Minimum	18	1	0
25 %	27	2,79	0
50 %	35	6,80	1
75 %	42	14,01	1
Maximum	70	60	1

The detailed classification reports show a consistent pattern in both models. Class “1” (presumably the preference for Sans Serif typography, given the summary statistics) has considerably higher precision, recall, and f1-score metrics compared to class “0”. This indicates that both models are more effective

identifying users who prefer Sans Serif typefaces but are less able to correctly identify those who do not (figure 1).

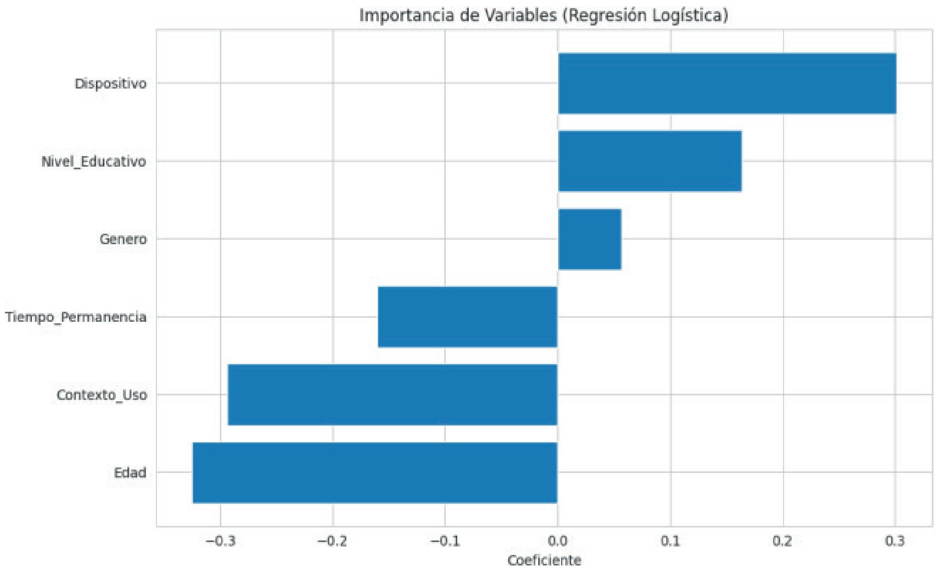

Figure 1. Importance of variables

Specifically, Logistic Regression achieves a recall of 0,89 for class “1”, while SVM achieves an identical value. However, the precision for class “0” is relatively low in both cases (Logistic Regression: 0,52, SVM: 0,48), implying a higher proportion of false positives for this class (figure 2).

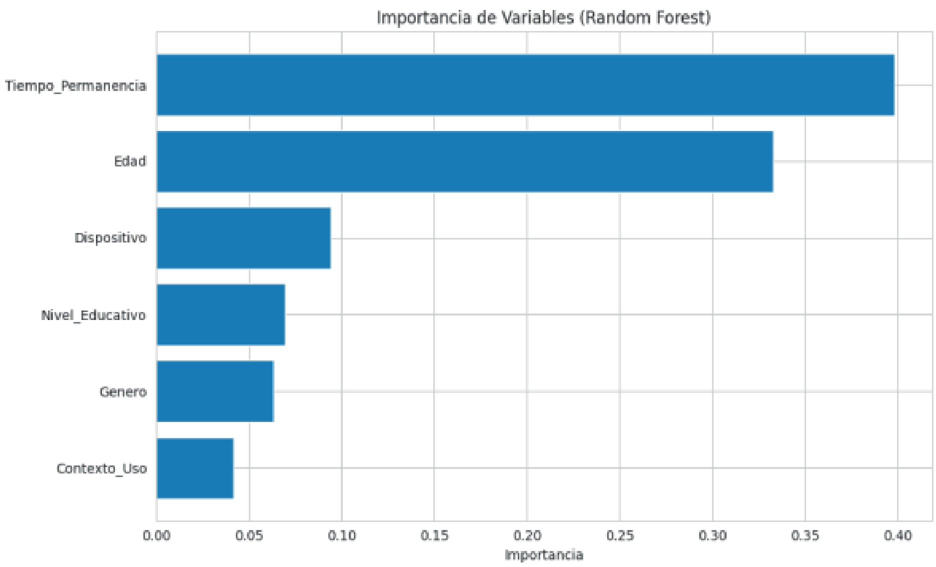

Figure 2. Importance of variables

The Random Forest shows similar cross-validation accuracy (0,631±0,059), but a closer analysis reveals lower precision and recall metrics performance for both classes, resulting in a lower overall accuracy (0,57). Similarly, the XGBoost model has the lowest cross-validation accuracy (0,590±0,043) and the most unfavorable classification metrics, as seen in figure 3.

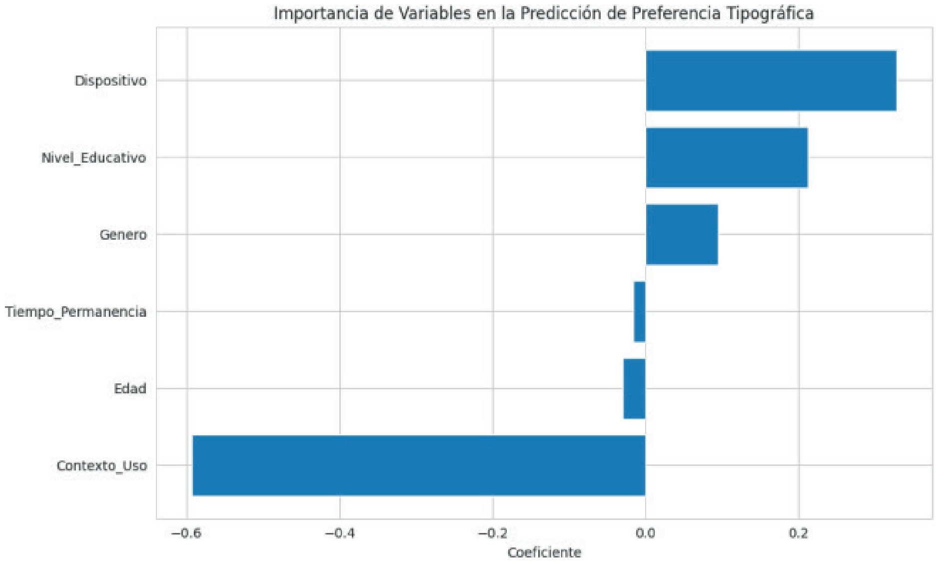

Figure 3. Importance of variables in the prediction of typographic preference

The performance evaluation through the ROC curve and the Area Under the Curve (AUC) calculation complements these findings. Logistic Regression obtains the highest AUC value (0,668), followed by SVM (0,608), suggesting a better ability to discriminate between classes compared to Random Forest (0,532) and XGBoost (0,541). An AUC of 0,5 would indicate random classification, so the values obtained suggest a modest discriminative ability, with Logistic Regression being slightly superior in this respect (figure 4).

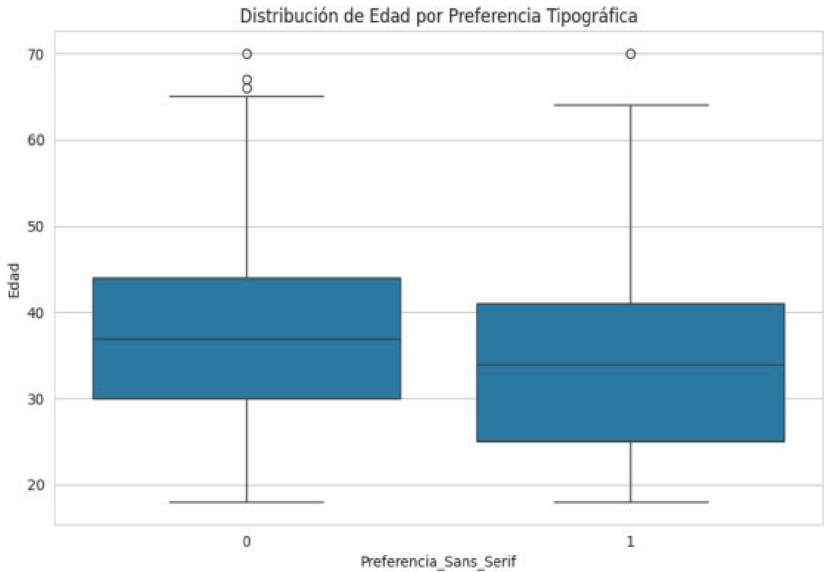

Figure 4. Age distribution by typographic preference

The summary statistics for the variables "Age", "Time_Permanence" and "Sans_Serif_Preference" provide descriptive information about the dataset used to train the models. The mean of "Sans_Serif_Preference" of 0,633 indicates a prevalence of users preferring

This typeface in the sample as indicated in table 2.

Tabla 2. Preference Sans Serif

Context of Use	Sans Preference Serif = 0	Preference Sans Serif = 1
Academic Reading	108	283
Recreational Reading	259	350

Finally, the distribution of preferences by use context reveals an interesting trend. In the context of

"Academic Reading," the preference for Sans Serif typefaces is notably higher (283) compared to Serif typefaces (108). In contrast, the difference is less pronounced in the "Recreational Reading" context (Sans Serif: 350, Serif: 259), although the preference for Sans Serif is still slightly higher. This contextual information could be crucial for refining future predictive models, possibly incorporating interactions between demographic and contextual variables.

In conclusion, Logistic Regression and SVM demonstrate slightly superior performance in the prediction task, albeit with limitations in identifying users who do not prefer Sans Serif typefaces. Contextual information on usage suggests that this variable plays a vital role in typographic preference and could be exploited to improve the accuracy of the models in future iterations.

DISCUSSION

The results obtained in this study reveal a moderate performance of the classification models in predicting typographic preferences, with Logistic Regression and SVM showing a slight superiority in accuracy (0,627 ± 0,052 and 0,634 ± 0,063, respectively). These findings are consistent with those who note that linear models such as Logistic Regression tend to be robust in binary classification tasks,[9] especially when the relationships between variables are moderately complex. However, the relatively low AUC (0,668 for Logistic Regression) suggests that the discriminative ability of the models is limited, which could be due to the subjective nature of typographical preferences influenced by psychological and contextual factors.[10,11]

A critical finding is a disparity in performance between classes: the models efficiently identify users who prefer Sans Serif typefaces (recall of 0,89) but fail to recognize those who do not (accuracy ≤ 0,52). This could be related to the class imbalance in the data, where 63,3 % of the sample prefers Sans Serif, a phenomenon documented in previous studies on biases in predictive models.[12,13]

Furthermore, the low accuracy for the minority class (Serif) suggests that the models confound patterns in this group, possibly due to the influence of variables not considered, such as context of use or cultural factors. Contextual analysis reinforced this hypothesis: in academic reading environments, the preference for Sans Serif was significantly higher (283 vs. 108), while in recreational reading, the difference was reduced (350 vs. 259). This supports research such as those who found that the purpose of the text moderates typographic preferences, as Sans Serif is associated with legibility in formal settings. On the other hand, the smaller gap in recreational contexts may be because typography in these cases is chosen for emotional or aesthetic factors.[14,15]

The inferiority of Random Forest and XGBoost (AUC < 0,55) contradicts the notion that tree-based models automatically outperform linear models on complex problems. One explanation is that these algorithms may be overfitting noise in the data, especially if the predictor variables (such as Age and Time Spent) have a weak non-linear relationship with the type preference.

It has a non-linear relationship with typographic preference. This underlines the importance of optimizing hyperparameters and considering interactions between variables, as suggested in the innovative typography model.[16,17]

The results highlight the need to incorporate contextual and psychological variables in predictive typography models, as factors such as the user's emotional state or the purpose of the text can be determinants.[18,19] In addition, data balancing techniques or ensemble approaches could improve performance in the minority class.[20] Finally, complementary qualitative studies would help to understand why users associate certain typefaces with specific contexts online about visual literacy in digital media.[21]

CONCLUSIONS

In conclusion, although linear models showed acceptable performance, their predictive ability is limited by the multifactorial complexity of typographic preferences.

Future research should integrate psychometric data and experimental designs that control variables like visual fatigue or cognitive load to move towards more accurate personalized recommendations.

BIBLIOGRAPHICAL REFERENCES

1. Bai Y, Huang Z, Gao W, Yang S, Liu J. Intelligent Artistic Typography: A Comprehensive Review of Artistic Text Design and Generation. APSIPA Trans Signal Inf Process. 2024;13(1). doi: http://10.48550/arXiv.2407.14774.

2. Iluz S, Vinker Y, Hertz A, Berio D, Cohen-Or D, Shamir A. Word-as-image for semantic typography. ACM Trans Graph. 2023;42(4):1-11. doi: http://10.48550/arXiv.2303.01818.

3. Günay M. The Impact of Typography in Graphic Design. Int J Eurasia Soc Sci. 2024;15(57):1446-64. doi: http://10.35826/ijoess.4519

4. Tanveer M, Wang Y, Mahdavi-Amiri A, Zhang H. Ds-fusion: Artistic typography via discriminated and stylized

diffusion. In: Proc IEEE/CVF Int Conf Comput Vis. 2023:374-84. doi: http://10.48550/arXiv.2303.09604.

5. Kim S, Jung AR, Kim Y. The effects of typefaces on ad effectiveness considering psychological perception and perceived communicator's power. J Mark Commun. 2021;27(7):716-41. doi: http://10.15444/GMC2018.08.08.05.

6. Brown NB. The Cognitive Type Project--Mapping Typography to Cognition. arXiv. 2024. doi: http://10.48550/arXiv.2403.04087.

7. Ezzeldeen O. The Abstract Character of Typography Art. J Des Sci Appl Arts. 2024;5(2):40-8. doi: http://10.21608/jdsaa.2024.229906.1365.

8. Poon ST. Typography design's new trajectory towards visual literacy for digital mediums. Stud Media Commun. 2021;9(1):9. doi: http://10.11114/smc.v9i1.5071.

9. Haenschen K, Tamul DJ, Collier JR. Font matters: Understanding typeface selection by political campaigns. Int J Commun. 2021;15:21. Available from: https://ijoc.org/index.php/ijoc/article/view/17615.

10. Rosidah S, Ivan FX, Murnani S, Nurlatifa H, Nugraha KA, Wibirama S. Impact of Color, Shape, and Typeface on Visual Attention: An Eye Tracking Study on Brand Logo. ASEAN J Sci Technol Dev. 2025;42(1):1-17. doi: http://10.29037/2224-9028.1608.

11. Rosyid HA, Putra AYH, Akbar MI, Dwiyanto FA. Can Multinomial Logistic Regression Predicts Research Group using Text Input?. Knowl Eng Data Sci. 2022;5(2):150-9. doi: http://10.17977/um018v5i22022p150-159.

12. Zhang M, Teng L, Xie C, Wang X, Foti L. Serif or sans serif typefaces? The effects of typefaces on consumers' perceptions of activity and potency of brand logos. Eur J Mark. 2025;59(4):879-922. doi: http://10.1108/EJM-06-2023-0497.

13. Vecino S, Gonzalez-Rodriguez M, Fernandez-Lanvin D, de Andres J. The impact of serif vs sans-serif typefaces on e-commerce websites. Int J Hum-Comput Interact. 2025;41(5):3613-24. doi: http://10.1080/10447318.2024.2338667.

14. Vecino S, Mehtali J, de Andrés J, Gonzalez-Rodriguez M, Fernandez-Lanvin D. How does serif vs sans serif typeface impact the usability of e-commerce websites?. PeerJ Comput Sci. 2022;8:e1139. doi: http://10.7717/peerj-cs.1139.

15. Minakata K, Beier S. The dispute about sans serif versus serif fonts: An interaction between the variables of serif and stroke contrast. Acta Psychol. 2022;228:103623. doi: http://10.1016/j.actpsy.2022.103623.

16. Mushtaq M, Ibrahim A, Adeel M, Murtaza G, Waleed A. Experimental Typography Typeface Design Inspired from Dates Palm Tree Bark Texture. Eur J Arts Humanit Soc Sci. 2025;2(1):54-9. doi: http://10.59324/ejahss.2025.2(1).06.

17. Chen W, Yang J, Wang Y. The influence of Chinese typography on information dissemination in graphic design: based on eye-tracking data. Sci Rep. 2024;14(1):13947. doi: http://10.1038/s41598-024-64964-y.

18. Tatsukawa Y, Shen IC, Qi A, Koyama Y, Igarashi T, Shamir A. FontCLIP: A Semantic Typography Visual-Language Model for Multilingual Font Applications. Comput Graph Forum. 2024;43(2):e15043. doi: http://10.1111/cgf.15043.

19. Cheng H, Xiao E, Gu J, Yang L, Duan J, Zhang J, et al. Unveiling typographic deceptions: Insights of the typographic vulnerability in large vision-language models. In: Eur Conf Comput Vis. Cham: Springer; 2024:179-96. doi:10.1007/978-3-031-73202-7_11.

20. Allal Z, Noura HN, Salman O, Chahine K. Leveraging the power of machine learning and data balancing techniques to evaluate stability in smart grids. Eng Appl Artif Intell. 2024;133:108304. doi: http://10.1016/j.engappai.2024.108304.

21. Mosleh MA, Gumaei AH. An Efficient Bidirectional Android Translation Prototype for Yemeni Sign

Language Using Fuzzy logic and CNN Transfer Learning Models. IEEE Access. 2024. doi: http://10.1109/ACCESS.2024.3512455.

FUNDING
The authors received no funding for the conduct of this research.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION
Conceptualization: Jazmín Isabel García-Guerra.
Data curation: Jazmín Isabel García-Guerra.
Formal analysis: **Héctor Aguilar-Cajas.**
Research: **Héctor Aguilar-Cajas.**
Methodology: **Héctor Aguilar-Cajas.**
Project management: **Héctor Aguilar-Cajas.**
Resources: Ana Lucía Rivera-Abarca.
Software: Ana Lucía Rivera-Abarca.
Supervision: Ana Lucía Rivera-Abarca.
Validation: Jazmín Isabel García-Guerra.
Visualization: Jazmín Isabel García-Guerra.
Writing - original draft: Ana Lucía Rivera-Abarca. Jazmín Isabel García-Guerra. Héctor Oswaldo Aguilar-Cajas. Heidy Elizabeth Vergara-Zurita. José Israel López-Pumalema. Freddy Armijos-Arcos.
Writing - review and editing: Ana Lucía Rivera-Abarca. Jazmín Isabel García-Guerra. Héctor Oswaldo Aguilar-Cajas. Heidy Elizabeth Vergara-Zurita. José Israel López-Pumalema. Freddy Armijos-Arcos.