

CERTIFICATE OF PARTICIPATION

This is to certify that
Salah Khennouf Salah
presented the Poster face-to-face

Towards Robust Arabic Authorship Attribution: A Transformer-Based Model for Multi-Author Imbalanced Corpora

at

“Computer Science and Technology”

Session during the **9th International Conference of Mathematical Sciences (ICMS 2025)**

held in Maltepe University, İstanbul Turkey on 03- 07 September 2025



Prof. Dr. Huseyin CAKALLI
Chairman of the Organizing Committee



Towards Robust Arabic Authorship Attribution: A Transformer-Based Model for Multi-Author Imbalanced Corpora

Salah Khennouf, Hassina Hadjadj, Mounir Bouras

Department of Electronics, Faculty of Technology, Mohamed Boudiaf University of M'sila (28000) Algeria, Speech Communication and Signal Processing Laboratory (LCPTS-USTHB)

Department of Telecommunication Faculty of Electrical Engineering, University of Science and Technology Houari Boumediene (USTHB) Bab Ezzouar, (16111) Algiers - Algeria

Department of Electronics, Faculty of Technology, Signal and Systems Analysis Laboratory (LASS) Mohamed Boudiaf University of M'sila (28000) Algeria

Abstract

Most conventional approaches to Authorship Attribution (AA) rely on statistical methods and classification algorithms that utilize stylistic features extracted from textual document. These features may include lexical, syntactic, structural, and content-based markers that reflect an author's unique writing style. In recent years, Pre-trained Language Models (PLMs) have gained significant attention in the field of text classification. While they have demonstrated strong performance on large-scale datasets and short texts, their effectiveness in scenarios with limited data—particularly in the context of Arabic Authorship Attribution (AAA)—remains insufficiently explored. This study aims to evaluate the effectiveness of state-of-the-art Pre-trained Arabic Transformer-based Models in handling imbalanced textual datasets, with a particular focus on the under-represented domain of theological law, which has witnessed limited contributions in the context of AA. The study addresses AAA using imbalanced corpora containing texts of varying lengths, extracted from several books written by Arab authors who lived during the same historical period. We conducted several experiments involving the fine-tuning of four Pre-trained Arabic Transformer-based Models: AraBERT, AraELECTRA, ARBERT, and MARBERT. The obtained experimental results have shown that AraBERT achieved the best performance in attributing texts to their respective authors.

Keywords: Deep Learning; Arabic transformer; AraBERT.

2020 Mathematics Subject Classification Numbers: 68T05, 68T07.

References

- [1] Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, 60(3), 538-556.
- [2] Zheng, W., Jin, M. (2023). A review on authorship attribution in text mining. *Wiley Interdisciplinary Reviews: Computational Statistics*, 15(2), e1584.