Overview of the Multi-Author Writing Style Analysis Task at PAN 2026

Eva Zangerle[1], Maximilian Mayerl[2], Martin Potthast[3] and Benno Stein[4]

[1]University of Innsbruck
[2]University of Applied Sciences BFI Vienna
[3]University of Kassel, hessian.AI, and ScaDS.AI
[4]Bauhaus-Universität Weimar

pan@webis.de https://pan.webis.de

Abstract

The multi-author writing style analysis task at PAN 2026 challenges participants to detect authorship changes between consecutive sentences within a document. Accurately identifying these positions is a fundamental step for downstream applications such as authorship attribution, plagiarism detection, or the identification of gift authorship. To assess the robustness of the submitted approaches under varying conditions, the participants had to detect authorship changes at sentence-level given three datasets of increasing difficulty. The datasets have been constructed by systematically controlling both the topical and the stylistic similarity between consecutive sentences. This paper provides an overview of the task, describes the dataset and its construction, summarizes the approaches submitted by participants, and discusses the evaluation results.

1. Introduction

The analysis of multi-author documents aims to identify stylistic changes within a text that may indicate a change in authorship. This task relies solely on intrinsic, stylistic evidence found in the document itself, without requiring external reference corpora or external sources. Detecting style changes is a fundamental task in authorship analysis and supports a range of downstream applications, including authorship attribution, plagiarism detection, forensic linguistics, and the analysis of collaboratively authored documents.

The multi-author writing style analysis task has constantly evolved since 2016. Initially, the task was framed as a style change detection task. Participants were asked to segment documents by author and cluster segments written by the same author [1]. In the 2017–2019 editions, the core task was to classify whether a document had been written by a single or by multiple authors [2, 3, 4]. In 2020, the task changed to determining the number of authors for a given document [5]. In the following years, 2021–2024, participants were challenged to identify all style and authorship changes between pairs of consecutive paragraphs [6, 7, 8, 9]. Driven by progress made predominantly through the use of pretrained language models, the task was refined again in 2025 by moving from paragraph-level to sentence-level analysis [10].

The methods developed and applied by participants have continuously co-evolved alongside the task. In the earlier editions, participants relied on extracting established stylometric signals [11], including lexical, syntactic, and structural features. The resulting stylometric profiles describing parts of the document were then fed into a classifier for binary classification to determine whether a style change had occurred. However, in recent years, approaches have been dominated by large language models that are fine-tuned on the (augmented) training set [12, 13, 14, 15, 16, 17].

This year, we continue the sentence-level style and authorship detection task while introducing new datasets based on fan fiction stories collected from the Archive of Our Own[1] platform. Drawing on these stories, we created three datasets by systematically managing topical and stylistic similarity

CLEF 2026 Working Notes, 21 – 24 September 2026, Jena, Germany

[1]https://archiveofourown.org/

during document generation, yielding an easy dataset in which topical cues are informative and stylistic cues are not the sole signal. In contrast, for the hard dataset, the topic is very homogeneous, forcing participants to leverage stylistic cues in their approaches.

Section 2 presents the task, the datasets, and the evaluation procedure. Section 3 provides a summary of the approaches submitted to the task, and Section 4 presents and discusses the results of the participating teams.

2. Style Change Detection Task

In the following, we present an overview of the task, the datasets employed, and the evaluation setup.

2.1. Task Definition

The core task of the multi-author writing style analysis task at PAN 2026 is to detect writing style changes within documents at the sentence level. For each pair of consecutive sentences in a document, the problem is formulated as a binary classification task that determines whether the two sentences were written by the same author or by different authors.

To systematically vary the difficulty of the task, three datasets of increasing complexity were provided. Participants were required to perform the classification task independently on each of the following three datasets:

- **Easy.** Documents cover multiple topics, allowing participants to exploit topic shifts as a strong signal for detecting changes in authorship. In addition, adjacent sentences exhibit low stylistic similarity, making stylistic changes more pronounced.
- **Medium.** Documents are more topically coherent, which reduces the usefulness of topic shifts and requires participants to rely more heavily on stylistic features. The stylistic similarity between adjacent sentences is moderate.
- **Hard.** All sentences within a document concern the same topic and exhibit a high degree of stylistic similarity, effectively eliminating topic variation as a discriminative signal and making authorship changes substantially more difficult to detect.

2.2. Dataset

For 2026, we curated a novel dataset based on fan fiction stories. Fan fiction consists of stories written by fans based on existing works such as TV series, movies, novels, or manga.

We relied on fan fiction data crawled from the Archive of Our Own platform. From approximately 8 million fan fiction stories, we selected the 100 largest fandoms (communities of fans centered around specific works, e.g., Harry Potter) based on the number of stories they contained. Each story is manually tagged by its author with information such as main characters, relationships, or content warnings. To control for topical similarity, we grouped stories within each fandom by their main character pairs and retained the stories of the top 100 main character pairs. For instance, within the Harry Potter fandom, we extracted all stories featuring character pairs such as Harry Potter and Hermione Granger.

Based on the extracted stories from the top 100 fandoms, grouped by the 100 main character pairs within each fandom, we curated a collection of individual documents. First, we applied several preprocessing steps, such as verifying that each story was written in English and removing HTML tags. We then extracted individual sentences from fan fiction stories sharing the same character pairs, written by two to four authors. For each of the extracted sentences, we computed two sentence representations: a stylistic feature vector capturing a variety of stylistic features and a semantic feature vector derived from Sentence Transformers (SBERT) [18]. These two sentence representations allowed us to compute semantic (topical) and stylistic similarity between sentences, which is the core component for the mixing step. To construct the documents for the datasets for the multi-author writing style analysis task,

we then concatenated sentences based on their topical and stylistic similarity, allowing us to control for the difficulty of the style detection task. For the three datasets, we configured the similarity threshold between consecutive sentences to be (1) relatively low for the *Easy* dataset, (2) moderate for the *Medium* dataset, and (3) high for the *Hard* dataset. Each of the three datasets contains 15,000 documents.

2.3. Evaluation

For each of the three datasets, participants were provided with predefined training, validation, and test splits (70%, 15%, 15%). Participants were required to submit their code to the TIRA platform [19]. During the competition, the test sets were withheld to ensure a fair evaluation. System performance was assessed using the macro-averaged F_1-Measure computed across all documents. Furthermore, we collected resource utilization data for each of the approaches, such as GPU energy consumption and overall runtime.

3. Survey of Submissions

We received eleven valid software submissions and ten paper submissions, nine of which were accepted as 'working notes' papers. One accepted paper was not published because no camera-ready version was submitted. In the following, we provide a brief summary of the approaches submitted.

Bodnar et al. evaluated three different approaches individually: fine-tuned DistilBERT, BERT, and SimCSE (Simple Contrastive Learning of Sentence Embeddings), a framework for contrastive learning for sentence embeddings. After preliminary experiments, the authors employed BERT fine-tuned for the binary classification task.

Elsayed et al. [20] introduce StyleGate, a parameter-efficient dual-branch LoRA architecture built on a frozen DeBERTa-v3-large cross-encoder. Two separate LoRA adapters are trained: a topic branch on easy data, intended to capture topical shifts, and a higher-rank style branch on medium and hard data, intended to capture more fine-grained stylometric cues. After both branches are trained and frozen, a lightweight MLP gate learns to blend their logits for each sentence pair, effectively acting as a per-sample mixture-of-experts mechanism. Training is carried out in three phases, first learning the topic branch, then the style branch, and finally the gate over all difficulty levels.

Fawzy et al. [21] introduce ContraStyle, a content-agnostic ensemble designed to suppress topical shortcuts and emphasize stylistic signals. Its main component is a DeBERTa-v3 pairwise classifier trained on windowed sentence pairs with a composite loss combining cross-entropy, supervised contrastive learning, R-Drop, and a Content-Agnostic Contrastive Objective that uses same-document style-change pairs as hard negatives. This transformer branch is complemented by a LightGBM classifier over handcrafted stylometric and SBERT-distance features, as well as a sequential BiLSTM style profiler over frozen SBERT sentence embeddings. The three components are combined through per-difficulty weighted ensembling and threshold calibration, with larger context windows and content-agnostic losses especially targeting the hard dataset.

Gagala [22] relies on DeBERTa for the encoding of sentences. The last two layers of the encoder were fine-tuned for the classification task, together with the classification head. Two additional embeddings capturing text length and the data source were also used as input for the classification. The author also experimented with data augmentation by adding further training data from Wikipedia; however, this did not improve results.

Kamthankar and Yardi [23] present a four-stage ensemble pipeline combining metric learning, classifier fusion, and document-level sequential refinement. Three transformer encoders, DeBERTa-v3, RoBERTa-large, and MPNet, are first fine-tuned with a corrected CoSENT ranking objective to obtain style-sensitive sentence embeddings. For each encoder and difficulty level, MLP boundary classifiers are trained on sentence-pair features derived from the two embeddings, their difference, product, and cosine similarity, using Focal BCE and R-Drop regularization. The resulting predictions are fused with Optuna-optimized ensemble weights and then refined by a BiLSTM over the full boundary sequence of

each document, allowing the system to model transition consistency beyond independent sentence-pair decisions.

Khalasi [24] proposes a hybrid approach combining SBERT sentence embeddings and 12 stylometric features, with classification via a Gradient Boosting-based classifier. Notably, they show the differences in performance of the two components across the three datasets, with SBERT features failing on the hard dataset due to the low topical variance.

Khelfaoui et al. [25] propose a task-routed specialist architecture based on Ettin-400M cross-encoders. Instead of training a single model for all difficulty levels, they fine-tune one sentence-pair cross-encoder per regime, with difficulty-specific positive class weights and decision thresholds. A separate document-level Ettin router classifies each input document as easy, medium, or hard using a head-tail sample of the document, and dispatches all adjacent sentence pairs to the corresponding specialist. This design explicitly treats the three tracks as different content-style regimes, allowing the easy model to exploit topical shifts while the hard model focuses more strongly on style cues.

Škurla [26] also presents a hybrid approach in which DeBERTa is the basis for the transformer-based branch capturing contextual and semantic information, whereas stylometric features classified using LightGBM are used in the stylometric branch. The two branches are combined in a late fusion approach using logistic regression. In contrast to the findings by [24], for this approach contextual and semantic information contributed most to the performance on the hard dataset.

Voznyuk and Tammaa [27] frame style-change detection as windowed boundary classification. For the medium and hard tracks, each candidate boundary is represented as a two-segment DeBERTa-v3 input with two sentences of context on both sides, enriched with short stylometric tag sentences encoding sentence complexity, punctuation, and first-person style cues. The model combines a weighted binary cross-entropy objective with a supervised contrastive auxiliary loss over a projection head. The authors further tune the data recipe separately for medium and hard, using a DAPT+TAPT-adapted encoder for medium and cross-group augmentation with negative downsampling for hard, followed by threshold tuning and document-level author-count calibration.

4. Evaluation Results

The evaluation results for the multi-author writing style analysis task at PAN 2026 are shown in Table 1. The table gives the F_1 scores on each of the three datasets for all of the submitted approaches. The

Table 1
Overall results for the multi-author analysis task, ranked by average F_1 performance across all three datasets. Best results are marked in bold.

Team	F_1(easy dataset)	F_1(medium dataset)	F_1(hard dataset)
stylefusion [26]	**1.000**	0.767	0.775
aimoment [21]	**1.000**	0.741	0.773
ulille [25]	0.995	0.712	0.769
spcrc [23]	0.993	0.722	0.757
datateam-tug [27]	0.918	**0.768**	**0.780**
holovchyk-ind	0.972	0.699	0.733
avocadobeans [22]	0.918	0.617	0.664
rheaveauro [24]	0.997	0.558	0.580
stitze	0.999	0.536	0.441
narrators [20]	0.968	0.560	0.447
redblinkinglight	0.410	0.489	0.441
Baseline Predict 0	0.410	0.489	0.441
Baseline Predict 1	0.234	0.042	0.173

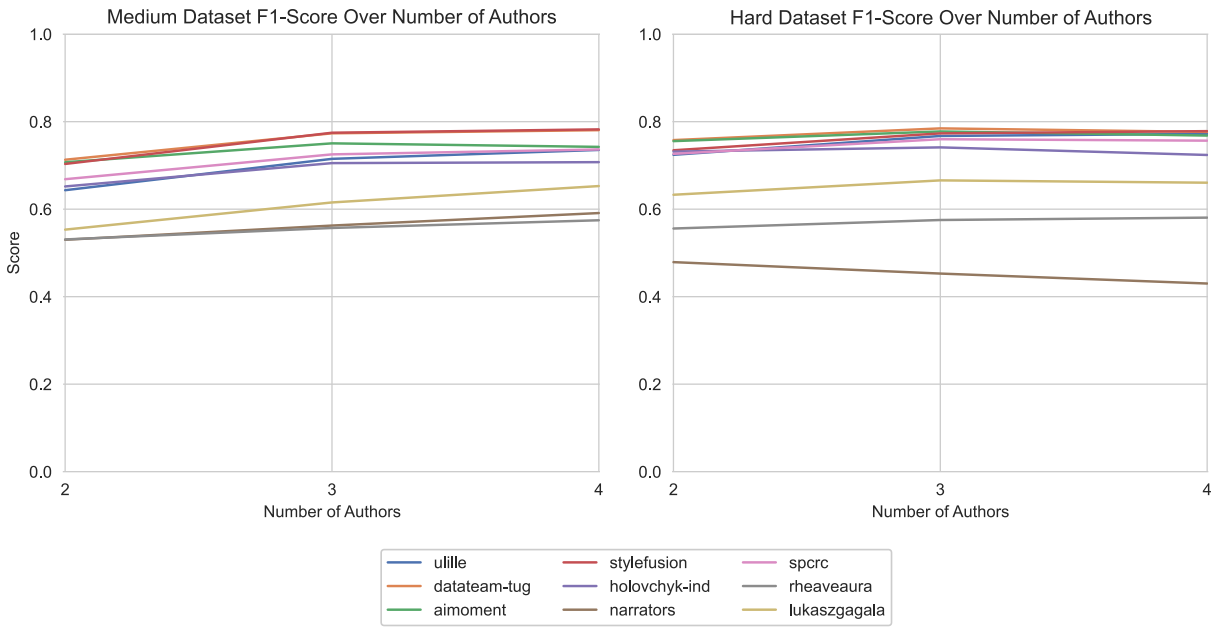

Figure 1: F_1 scores per team for the medium (left) and hard (right) datasets, depending on the number of authors per document.

best results for each dataset are highlighted in bold. Also shown are the results for two simple baseline models; one baseline predicts a style change on every sentence boundary (Predict 1), while the other baseline never predicts a style change (Predict 0).

For all three datasets, all of the submitted approaches outperform the baseline, in most cases by a significant margin. For the easy dataset, two approaches (Fawzy et al. [21] and Škurla [26]) achieved a perfect F_1 score of 1. All other approaches also achieved very high scores for this dataset, with the smallest score being 0.918. This is a change from last year, when only about half of the submitted approaches managed to score above 0.9 on the easy dataset. We can therefore conclude that the easy version of the task, where a strong topic signal exists to facilitate the detection of an authorship change, is essentially solved.

For the medium dataset, the best score of 0.768 was achieved by Voznyuk and Tammaa [27], with Škurla [26] coming in as a close second with a score of 0.767. These scores are lower than the best medium score of last year, which was 0.825. Since the average score for the medium dataset in general has gone down slightly, this is most likely because the dataset used this year is overall more difficult.

For the hard dataset, the best F_1 score, 0.780, was achieved by Voznyuk and Tammaa [27], which is the same approach that also achieved the best score on the medium dataset. Interestingly, this is a higher score than the one achieved on the medium dataset. This is in line with what we observed in last year's edition of the task: since the medium dataset mostly controls for topic, making topics largely homogeneous across a generated document while still including some topic variation, this might suggest that models have difficulty dealing with documents in which both topical and stylometric signals are important for detecting an authorship change. Similar to the medium dataset, the scores achieved this year on the hard dataset are also, on average, slightly lower than last year.

As in previous years, we also looked at the performance of each approach depending on the number of authors in a document. The results for this, on the medium and hard datasets, are shown in Figure 1. For this year's approaches, we observe the peak in performance at four authors (for medium) and three authors (for hard). Almost all of the submitted approaches performed worse on documents with only two distinct authors.

Finally, we also looked at the resource utilization of the submitted approaches. This seems particularly relevant this year, given how the cost of computing resources has increased rapidly over the last months. The overall results are shown in Table 2. A visualization, showing the achieved score on the hard dataset

Table 2
Performance metrics for the approaches that were run on the hard dataset. For CPU, GPU, RAM and VRAM, the values are the maximum utilization / usage during its execution. For GPU energy consumption, values are given in joules. Finally, time is given in milliseconds.

Team	F_1(hard dataset)	Time	CPU	GPU	RAM	VRAM	GPU Energy
ulille	0.769	1520	19	100	25350	16063	227336
datateam-tug	**0.780**	1140	10	100	25370	39306	704278
aimoment	0.773	1740	19	100	25612	15998	583749
stylefusion	0.775	10780	20	100	25657	41445	1270490
holovchyk-ind	0.733	4700	20	100	25354	39516	1716859
narrators	0.447	20800	20	100	25333	39394	1701366
spcrc	0.757	3300	232	83	25608	4854	335074
rheaveaura	0.580	70750	20	100	25210	42670	342626
avocadobeans	0.664	7870	250	100	25661	40908	1083290

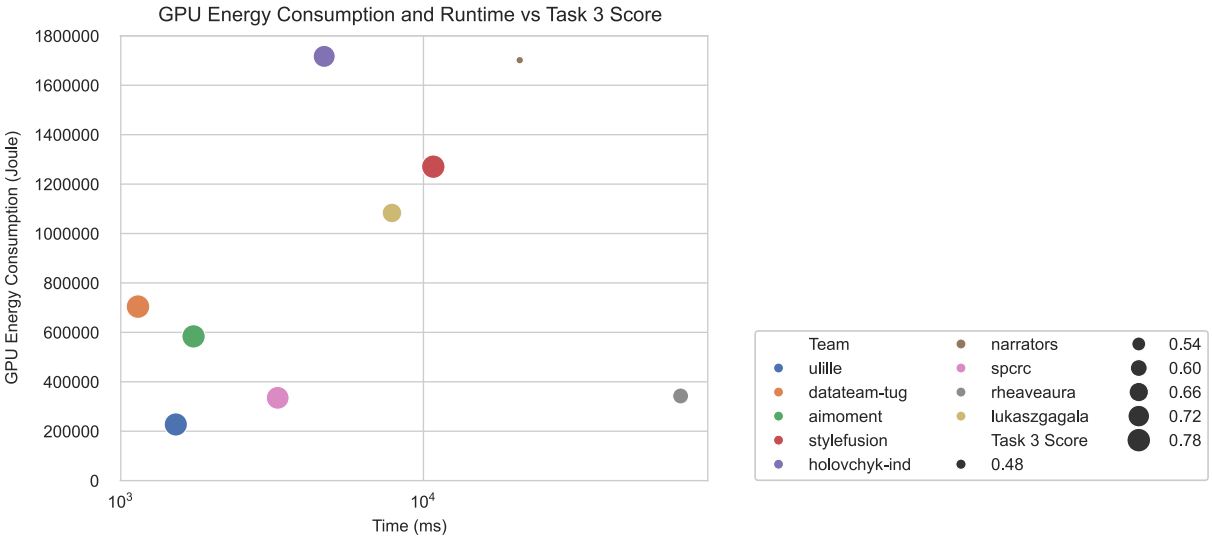

Figure 2: GPU energy consumption and runtime versus F_1 score on the hard dataset for each team. Time is shown in log-scale.

versus GPU energy consumption and overall runtime is shown in Figure 2. The F_1 score is visualized via the size of the dots. One particular approach stands out here: Khelfaoui et al. [25] is only slightly behind other approaches in terms of the achieved F_1 score (being behind by 0.011 compared to the highest score), but has the second-lowest runtime and the lowest GPU energy consumption among all submitted approaches.

5. Conclusion

For the 2026 edition of the multi-author writing style analysis task at PAN, we once again asked participants to identify whether a change in authorship occurs at sentence boundaries. Unlike in the past few editions of the task, the datasets used this year were based on fan fiction of popular media franchises, i.e. narrative content, as opposed to the more discussion-based content used in previous years. We once again presented the task in three levels of difficulty, related to the extent at which topic variations were present within a document, allowing approaches to exploit topic information on the easy difficulty dataset.

For the first time this year, all submitted approaches achieved very high scores on the easy dataset. The easy version of the task could therefore be considered solved. For the medium and hard difficulty

levels, we saw a slight decrease in performance compared to last year, suggesting that it may be more difficult to distinguish authors based purely on writing style in narrative content than it is in discussion-based content.

Declaration on Generative AI

The authors used ChatGPT and DeepL for grammar and spelling checks as well as for paraphrasing. After using these services, the authors reviewed the content and revised it as necessary.

References

[1] E. Stamatatos, M. Tschuggnall, B. Verhoeven, W. Daelemans, G. Specht, B. Stein, M. Potthast, Clustering by Authorship Within and Across Documents, in: Working Notes Papers of the CLEF 2016 Evaluation Labs, CEUR Workshop Proceedings, CLEF and CEUR-WS.org, 2016. URL: http://ceur-ws.org/Vol-1609/.

[2] M. Kestemont, M. Tschuggnall, E. Stamatatos, W. Daelemans, G. Specht, B. Stein, M. Potthast, Overview of the Author Identification Task at PAN-2018: Cross-domain Authorship Attribution and Style Change Detection, in: Working Notes Papers of the CLEF 2018 Evaluation Labs, volume 2125 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018. URL: https://ceur-ws.org/Vol-2125/invited_paper_2.pdf.

[3] M. Tschuggnall, E. Stamatatos, B. Verhoeven, W. Daelemans, G. Specht, B. Stein, M. Potthast, Overview of the Author Identification Task at PAN 2017: Style Breach Detection and Author Clustering, in: Working Notes Papers of the CLEF 2017 Evaluation Labs, volume 1866 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017. URL: http://ceur-ws.org/Vol-1866/.

[4] E. Zangerle, M. Tschuggnall, G. Specht, M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2019, in: CLEF 2019 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2019. URL: http://ceur-ws.org/Vol-2380/paper_243.pdf.

[5] E. Zangerle, M. Mayerl, G. Specht, M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2020, in: CLEF 2020 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2020. URL: https://ceur-ws.org/Vol-2696/paper_256.pdf.

[6] E. Zangerle, M. Mayerl, , M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2021, in: CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2021, pp. 1760–1771. URL: https://ceur-ws.org/Vol-2936/paper-148.pdf.

[7] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2022, in: CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2022. URL: http://ceur-ws.org/Vol-3180/paper-186.pdf.

[8] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Multi-Author Writing Style Analysis Task at PAN 2023, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), volume 3497 of *CEUR Workshop Proceedings*, 2023, pp. 2513–2522. URL: https://ceur-ws.org/Vol-3497/paper-201.pdf.

[9] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Multi-Author Writing Style Analysis Task at PAN 2024, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2024.

[10] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Multi-Author Writing Style Analysis Task at PAN 2025, in: Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2025, Madrid, Spain, 9-12 September 2025, volume 4038 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, pp. 3568–3574. URL: https://ceur-ws.org/Vol-4038/paper_279.pdf.

[11] E. Stamatatos, Intrinsic Plagiarism Detection Using Character n-gram Profiles, in: SEPLN 2009 Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 09), Universidad Politécnica de Valencia and CEUR-WS.org, 2009, pp. 38–46. URL: http://ceur-ws.org/Vol-502.

[12] A. Iyer, S. Vosoughi, Style Change Detection Using BERT—Notebook for PAN at CLEF 2020, in: CLEF 2020 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2020. URL: http://ceur-ws.org/Vol-2696/.

[13] Z. Zhang, Z. Han, L. Kong, X. Miao, Z. Peng, J. Zeng, H. Cao, J. Zhang, Z. Xiao, X. Peng, Style Change Detection Based On Writing Style Similarity—Notebook for PAN at CLEF 2021, in: CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2936/paper-198.pdf.

[14] T.-M. Lin, C.-Y. Chen, Y.-W. Tzeng, L.-H. Lee, Ensemble Pre-trained Transformer Models for Writing Style Change Detection, in: CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2022. URL: http://ceur-ws.org/Vol-3180/paper-210.pdf.

[15] H. Chen, Z. Han, Z. Li, Y. Han, A Writing Style Embedding Based on Contrastive Learning for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2023, pp. 2562–2567. URL: https://ceur-ws.org/Vol-3497/paper-206.pdf.

[16] J. Lv, Y. Yi, H. Qi, Team Fosu-stu at PAN: Supervised fine-tuning of large language models for Multi Author Writing Style Analysis, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2024.

[17] X. Lin, Z. Han, C. Liu, X. Duan, Style Change Detection in Multi-Author Writing: A Deep Learning Approach Based on DeBERTa, in: Working Notes of CLEF 2025 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2025.

[18] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using Siamese BERT-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992. URL: https://aclanthology.org/D19-1410/. doi:10.18653/v1/D19-1410.

[19] M. Fröbe, M. Wiegmann, N. Kolyada, B. Grahm, T. Elstner, F. Loebe, M. Hagen, B. Stein, M. Potthast, Continuous Integration for Reproducible Shared Tasks with TIRA.io, in: Advances in Information Retrieval. 45th European Conference on IR Research (ECIR 2023), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2023, pp. 236–241.

[20] A. Elsayed, E. Mohammed, N. El-Makky, M. Torki, Team narrators at PAN 2026: StyleGate: Dual-Branch LoRA on DeBERTa-v3 for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2026.

[21] M. Fawzy, N. El-Makky, M. Torki, Team aimoment at PAN 2026: Content-Agnostic Contrastive Stylometry for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2026.

[22] L. Gagala, Team avocadobeans at PAN 2026: Style Change Detection with DeBERTa, Evidence Aggregation and Sentence Length Embedding, in: Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2026.

[23] S. Kamthankar, A. Yardi, Team SPCRC at PAN 2026: Multi-Encoder Ensemble with Sequential BiLSTM Refinement for Style Change Detection, in: Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2026.

[24] P. Khalasi, Team rheaveaura at PAN 2026: Stylometric-Enhanced Sentence Embeddings for Multi-Author Writing Style Change Detection, in: Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2026.

[25] A. Khelfaoui, D. Sileo, G. Loiseau, Team ULILLE at PAN 2026: Task-Routed Ettin Cross-Encoders for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2026.

[26] A. Škurla, Team StyleFusion at PAN 2026: DeBERTa Meets Stylometry, in: Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2026.

[27] A. Voznyuk, A. Tammaa, Contrastive Windowed Boundary Classification for Multi-Author Writing Style Analysis, in: Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2026.