Team SPCRC at PAN 2026: Multi-Encoder Ensemble with Sequential BiLSTM Refinement for Style Change Detection

Notebook for PAN at CLEF 2026

Shubhankar Kamthankar[1], Arti Yardi[1]

[1] International Institute of Information Technology Hyderabad (IIIT-H), Gachibowli, Hyderabad 500032, India

Abstract

We present a multi-stage style change detection system for the PAN 2026 Multi-Author Writing Style Analysis shared task, which requires identifying sentence-level authorship transitions within documents. The proposed pipeline combines heterogeneous transformer encoders, ensemble fusion, and document-level sequential refinement. Three sentence encoders (DeBERTa-v3-base, RoBERTa-large, and MPNet-base) are first fine-tuned using the CoSENT ranking objective to learn style-sensitive sentence representations. Their predictions are subsequently combined through difficulty-specific MLP boundary classifiers trained with Focal BCE loss and R-Drop regularization to address severe class imbalance. The resulting probabilities are fused using Optuna-based ensemble optimization and further refined using a bidirectional LSTM operating over complete document boundary sequences. Compared to earlier paragraph-level formulations of the task, the sentence-level granularity of PAN 2026 produces substantially longer transition sequences per document, making sequential modeling considerably more informative. On the official hidden test set evaluated through TIRA, the submitted system achieves macro-F1 scores of **0.9930**, **0.7218**, and **0.7573** on the easy, medium, and hard subsets respectively, while maintaining close agreement with validation performance across all difficulty levels.

Keywords

Style Change Detection, Authorship Analysis, Sentence Embeddings, CoSENT, Ensemble Learning, BiLSTM, PAN 2026

1. Introduction

Style change detection (SCD) asks a system to identify, within a single document, the positions at which authorship changes from one writer to another. Since its introduction as a PAN shared task in 2017 [1, 2], the problem has evolved through several reformulations: from coarse document-level decisions, through paragraph-level boundary classification, to the sentence-level formulation used in PAN 2026. Each shift in granularity fundamentally changes the modelling assumptions that are appropriate.

Earlier paragraph-level systems achieved strong results by treating each consecutive paragraph pair independently: a single classifier trained on embedding-space features of two paragraphs was often sufficient [3, 4]. Sentence-level SCD breaks this assumption in two ways. First, representations must encode style efficiently at short-text granularity, since any individual sentence is too short to reliably characterise an author's style in isolation. Second, predictions for neighbouring boundaries are no longer independent — the sequence of boundary probabilities across a document carries structural information about plausible author-transition patterns that a per-pair classifier cannot access.

We address both challenges with a four-stage pipeline. For representation quality, we fine-tune three complementary transformer encoders with the CoSENT ranking loss [5], which directly optimises cosine-similarity ordering between same- and different-author pairs — a tighter training signal for embedding-based boundary detection than standard cross-entropy or contrastive objectives. For sequence structure, we introduce a bidirectional LSTM that reads the full document boundary sequence and refines each pairwise prediction in context.

CLEF 2026 Working Notes, September 2026, Jena, Germany

*Corresponding author.

Contributions

1. We show that initialising CoSENT fine-tuning from NLI- and SBERT-pretrained checkpoints avoids cold-start instability and produces well-calibrated cosine similarities within a few epochs, reaching useful stylistic separation without the cold-start instability typical of randomly initialised metric learning.

2. We design difficulty-adaptive loss configurations that address the qualitatively different challenges of each subset: reduced focal γ and higher positive weighting for hard (correcting probability compression), and extreme positive weighting with a consecutive-change penalty for the severely imbalanced medium subset.

3. We show that a BiLSTM operating over full document boundary sequences provides consistent improvement over the flat ensemble, with the largest gains on medium (+0.026 F1) and hard (+0.013 F1) where local pairwise evidence is weakest.

4. We present a shard-based feature extraction pipeline that decouples encoder inference from sequential training, making large-scale BiLSTM refinement tractable without requiring all encoder representations in memory simultaneously.

2. Related Work

Style Change Detection at PAN. Early PAN SCD systems relied on handcrafted stylometric features — character and word n-gram distributions, punctuation profiles, average sentence length — combined with SVM or logistic regression classifiers [4]. The introduction of transformer-based encoders brought a shift towards learned representations, with systems using BERT and RoBERTa embeddings in pairwise classifiers [3]. Recent systems have explored contrastive learning [6] and multi-task objectives to jointly model attribution and segmentation. Our system builds on the transformer-based pairwise classification paradigm but introduces two extensions: a ranking loss that directly optimises the embedding geometry for cosine-similarity-based classification, and a sequential model that corrects pairwise predictions using document-level context.

Metric Learning for Authorship. Contrastive and ranking losses directly shape the embedding space and have proven effective for authorship tasks. CoSENT [5] optimises a listwise objective over all same- versus different-author pairs in a batch. Unlike triplet loss, which samples only a single violating triple at a time, CoSENT penalises all constraint violations simultaneously through a logsumexp, providing denser gradient signal per batch. We adopt CoSENT over these alternatives because our pairwise classifiers operate directly on cosine-similarity-derived features, making the alignment between training loss and downstream usage tighter.

Sequential Models for Segmentation. Text segmentation has a long history of sequential modelling, from the classical TextTiling algorithm [7] to neural approaches using BiLSTMs [8] and Transformers. For our setting, a BiLSTM is preferable to a Transformer sequence model: the input sequence (projected encoder features and boundary probabilities) is already a rich compressed representation; document boundary sequences are short enough that attention does not have a clear advantage over recurrence; and the BiLSTM's inductive bias towards ordered local dependencies is well-matched to the positional structure of author-transition patterns.

3. Task and Data

3.1. Task Definition

Given a document with n sentences $\{s_1, \ldots, s_n\}$ written by two or more authors in sequence, the system must predict $\mathbf{y} \in \{0, 1\}^{n-1}$, where $y_i = 1$ denotes a style change between s_i and s_{i+1}. The evaluation metric is macro-F1 over the positive (change) and negative (no-change) classes.

3.2. Dataset Characteristics

Three difficulty subsets reflect qualitatively different modelling challenges, summarised in Table 1.

Table 1
Dataset characteristics per difficulty subset. Change rate = fraction of boundaries labeled positive.

Subset	Ch. Rate	Contrast	Challenge
Easy	≈30%	High	Straightforward
Medium	≈4%	Low	Severe imbalance
Hard	≈20%	Subtle	Fine-grained

The *easy* subset has high inter-author style contrast and a moderate change rate. The *medium* subset is the hardest imbalance challenge: with only $\approx 4.3\%$ positive boundaries, a trivial all-negative classifier achieves high accuracy but near-zero macro-F1. The *hard* subset has moderate change rates but requires detecting subtle within-genre stylistic differences.

4. System Overview

Our system follows a four-stage pipeline designed to progressively refine style change predictions from local sentence representations to document-level boundary decisions. Stage 1 fine-tunes transformer encoders using the CoSENT ranking objective to learn style-sensitive sentence embeddings. Stage 2 constructs sentence-pair features and trains difficulty-specific MLP classifiers. Stage 3 combines encoder predictions through weighted ensemble fusion. Stage 4 applies sequential BiLSTM refinement over complete document boundary sequences to model transition consistency and global boundary structure.

The staged design offers practical advantages beyond modularity. Encoder fine-tuning is decoupled from downstream classifier training, so expensive transformer inference runs only once per dataset split and is cached. The BiLSTM can be retrained or recalibrated without recomputing any encoder representations.

Figure 1 illustrates the pipeline and information flow between stages.

5. Stage 1: CoSENT Encoder Fine-Tuning

5.1. Choice of Base Models

We select three encoder models with complementary architectures and pre-training regimes, specifically choosing checkpoints that already produce meaningful cosine similarities before any fine-tuning. This is a deliberate departure from starting with vanilla language model weights: an embedding space that already reflects semantic proximity allows CoSENT to specialise for stylistic discrimination rather than rebuild geometric structure from scratch.

DeBERTa-v3 contributes disentangled attention mechanisms that improve contextual token interactions. RoBERTa-large provides higher representational capacity (355M parameters) and broad masked-language pre-training. MPNet introduces permutation-based pre-training that improves sentence-level

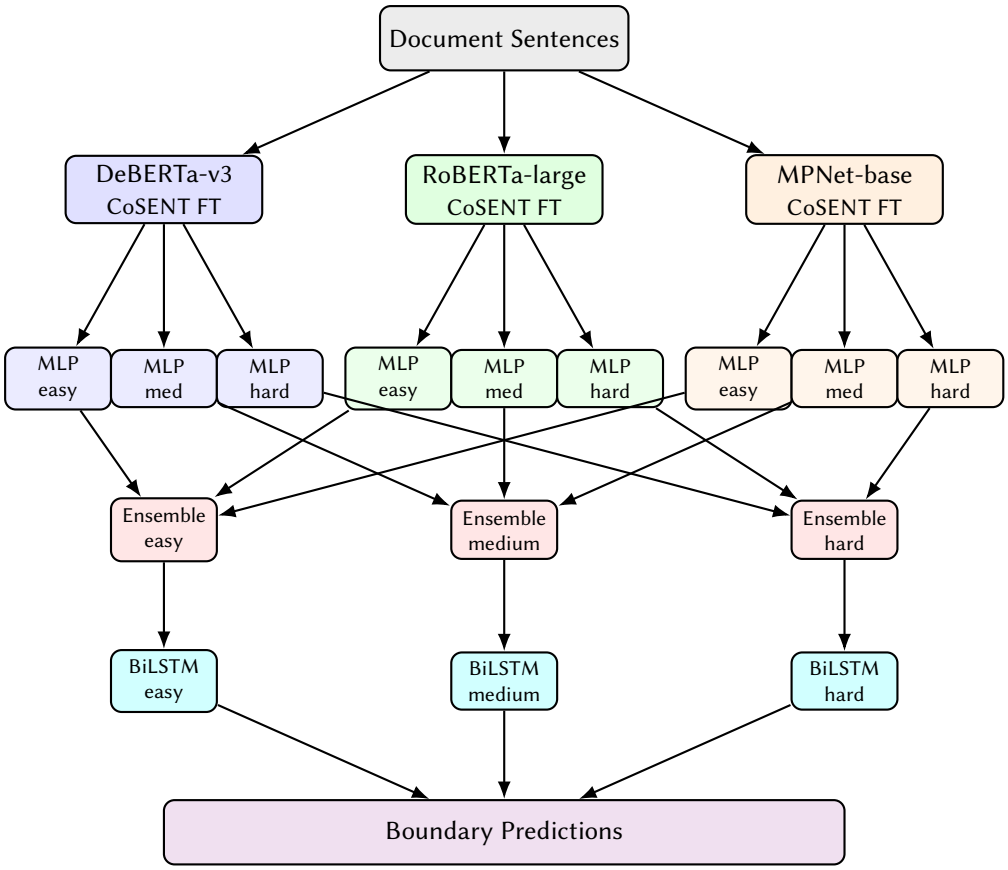

Figure 1: Complete inference pipeline showing all 9 MLP classifiers and 3 BiLSTMs explicitly. Colour encodes encoder origin: blue = DeBERTa, green = RoBERTa, orange = MPNet. Each difficulty column (easy / medium / hard) flows independently from classifier through ensemble to BiLSTM; at inference only the column matching the document's difficulty is active.

context aggregation. Together they form a heterogeneous ensemble whose diversity reduces correlated errors.

- **DeBERTa**: `cross-encoder/nli-deberta-v3-base` (768 d);
- **RoBERTa**: `all-roberta-large-v1` (1024 d);
- **MPNet**: `all-mpnet-base-v2` (768 d).

5.2. CoSENT Training Objective

Given a batch of sentence pairs with embeddings $(\mathbf{u}_i, \mathbf{v}_i)$ and label $\ell_i \in \{0, 1\}$, the CoSENT objective [5] is:

$$\mathcal{L}_{\text{CoSENT}} = \log\left(1 + \sum_{\substack{i:\,\ell_i=1 \\ j:\,\ell_j=0}} e^{\,\lambda(\cos\theta_j - \cos\theta_i)}\right), \tag{1}$$

where $\lambda{=}20$ is a temperature parameter and the sum ranges over all (positive pair i, negative pair j) combinations in the batch. The logsumexp formulation penalises all constraint violations simultaneously, providing denser gradient signal than triplet loss, which samples only a single violating triple at a time.

We note that the standard open-source CoSENT implementation contains a sign error in the pairwise difference term (penalising $\cos\theta_i - \cos\theta_j$ instead of the correct $\cos\theta_j - \cos\theta_i$), which inverts the ranking objective. Our implementation corrects this; the fix was critical for obtaining usable representations.

5.3. Pair Generation and Training

Positive pairs are consecutive same-author sentence pairs within documents; negative pairs span adjacent style-change boundaries. Group-aware sampling allows up to six within-group positives and three cross-boundary negatives per group, targeting a negative-to-positive ratio of 2:1 to avoid skewing the embedding space toward negative separation.

All three encoders are trained exclusively on the *hard* split. The hard split contains the subtlest stylistic transitions and minimises topical shortcuts: models cannot exploit lexical drift between topics to infer authorship, and must learn genuine stylistic discrimination. Representations learned here transfer across difficulty levels through the downstream classifiers.

Training uses mean-pooled token representations, bf16 mixed precision, gradient accumulation for an effective batch size of 128, learning rate 8×10^{-6}, cosine-with-warmup schedule (15% warmup), and early stopping with patience 5 over up to 12 epochs.

5.4. Training Dynamics

A consistent pattern emerged across all three encoders: CoSENT loss and embedding separation margin improved steadily during early epochs, but continued optimisation beyond a moderate separation point began to degrade downstream macro-F1. For DeBERTa, validation macro-F1 peaked at epoch 3 (0.7345) as the cosine margin reached 0.0819; further training pushed the margin to 0.11 but reduced F1 to 0.7164.

This suggests that overly aggressive metric separation collapses the within-class variance that short-text style classifiers depend on. Early stopping against downstream validation macro-F1 rather than the ranking loss itself was essential to avoid this regime.

6. Stage 2: Boundary Classification

6.1. Feature Construction

With encoders frozen, we construct a pairwise feature vector for each consecutive sentence pair (s_i, s_{i+1}) with embeddings $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$:

$$\mathbf{f} = \left[\mathbf{u};\ \mathbf{v};\ |\mathbf{u}-\mathbf{v}|;\ \mathbf{u}\odot\mathbf{v};\ \cos(\mathbf{u},\mathbf{v})\right] \in \mathbb{R}^{4d+1}. \tag{2}$$

The concatenation of embeddings, element-wise difference, and element-wise product follows the InferSent sentence-pair representation [9]. We additionally include explicit cosine similarity as a scalar, surfacing the dominant style-change signal directly rather than requiring the classifier to recover it from higher-dimensional projections. This also shortens the gradient path from the loss to the similarity computation, which benefits the BiLSTM's projection layers in Stage 4.

Nine separate classifiers are trained (three encoders $\times$ three difficulties), each an MLP with two hidden layers (512 and 256 units), LayerNorm, Tanh activations, and dropout rates of 0.3 and 0.2.

6.2. Difficulty-Adaptive Loss

All classifiers are trained with Focal BCE loss [10]:

$$\mathcal{L}_{\text{focal}} = -\frac{1}{N}\sum_i w^+(1-p_t)^\gamma \log p_t, \tag{3}$$

where γ down-weights easy examples and w^+ reweights the positive class. Table 2 gives the per-difficulty configuration with motivations.

The hard subset required particular calibration care. Initial experiments with γ=2.0, w^+=4.0 produced severely compressed probabilities: the optimal BiLSTM threshold dropped to 0.31, a clear signal that positive-class confidence was systematically suppressed. Reducing γ to 1.0 restores sensitivity to hard positives; raising w^+ to 6.0 provides the corresponding recall boost.

Table 2
Per-difficulty Focal BCE hyperparameters. τ_0 is the initial threshold before BiLSTM recalibration.

Subset	γ	w^+	τ_0	Key motivation
Easy	2.0	2.5	0.45	Mild imbalance; standard Focal
Medium	2.0	25.0	0.08	≈4% pos. rate; extreme upweighting needed
Hard	1.0	6.0	0.40	Lower γ prevents probability compression

R-Drop regularisation [11] (α=0.7) penalises KL divergence between two forward passes with different dropout masks, acting as implicit data augmentation. This was particularly useful for medium, where the positive class is so rare that standard regularisation alone allowed over-fitting to negative examples.

7. Stage 3: Ensemble Fusion

Predictions from the three encoder-specific classifiers are combined as:

$$\hat{p}_i^{\text{ens}} = \sum_k w_k \cdot \hat{p}_i^{(k)}, \tag{4}$$

where weights satisfy $0.05 \leq w_k \leq 0.65$ and $\sum_k w_k = 1$. The ensemble combines heterogeneous embedding spaces, reducing correlated errors that any individual encoder would make.

We use Optuna [12] to jointly search encoder weights and decision threshold τ over 300 trials per difficulty, maximising validation macro-F1. Joint optimisation of weights and threshold is important: an encoder whose probabilities are better calibrated at a particular threshold contributes differently than one that is better ranked but poorly calibrated, and optimising weights alone while fixing the threshold fails to capture this interaction. The resulting configuration is passed to Stage 4 as the ensemble prior.

8. Stage 4: Sequential BiLSTM Refinement

8.1. Motivation

The ensemble of Stages 1–3 assigns each boundary a probability independently of all others. This is suboptimal for two reasons. First, author-transition patterns are locally constrained: a style change at position i reduces the probability of another change at $i{+}1$ in most writing scenarios. Second, global document structure informs appropriate priors: a document with many authors will exhibit higher transition density than one with two, and a model that reads the full sequence can adjust its confidence accordingly.

8.2. Input Representation

For each boundary i in a document of n sentences, the BiLSTM input is:

$$\begin{aligned}\mathbf{x}_i = \big[&\text{Proj}_1(\mathbf{f}_i^{(1)});\ \ldots;\ \text{Proj}_K(\mathbf{f}_i^{(K)});\\ &\hat{p}_i^{(1)};\ \ldots;\ \hat{p}_i^{(K)};\ \hat{p}_i^{\text{ens}};\ \phi_i\big],\end{aligned} \tag{5}$$

where each $\text{Proj}_k(\cdot)$ is a Linear–LayerNorm–ReLU projection to 384 d, and $\phi_i \in \mathbb{R}^2$ contains two positional features: *relative position* $i/(n{-}2) \in [0,1]$ and *local change density* (mean ensemble

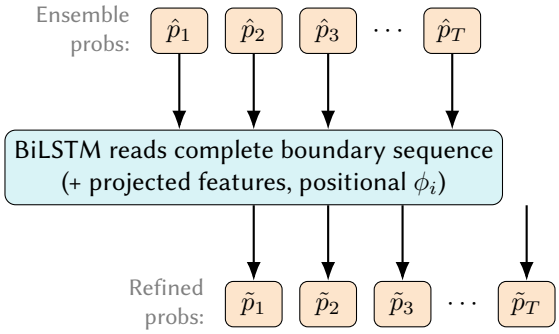

Figure 2: The BiLSTM reads the *entire* sequence of $T{=}n{-}1$ boundary representations in a single pass and emits a refined probability for every boundary simultaneously, capturing transition structure invisible to pairwise classifiers.

probability in a ± 2 boundary window). With $K{=}3$ encoders the total input dimension is $384{\times}3 + 3 + 1 + 2 = 1158$.

8.3. Architecture and Training

The BiLSTM has 2 layers, hidden size 256, and inter-layer dropout 0.3. The output head is Linear$(512{\to}64)$–ReLU–Dropout(0.2)–Linear$(64{\to}1)$, producing one logit per boundary.

Training follows the per-difficulty Focal BCE configuration of Stage 2. For the medium subset an additional consecutive-change penalty ($\lambda_{\text{cons}}{=}0.1$) penalises $\sum_i \hat{p}_i \cdot \hat{p}_{i+1}$, encoding a soft prior that adjacent boundaries are unlikely to both be style changes — a prior strongly supported by the 4% positive rate in that subset.

Thresholds are re-tuned after BiLSTM training via fine-grained grid search on the validation set, yielding $\tau_{\text{easy}}{=}0.52$, $\tau_{\text{medium}}{=}0.65$, $\tau_{\text{hard}}{=}0.57$ — substantially different from the Stage 3 thresholds, reflecting the BiLSTM's altered probability calibration.

9. Implementation and Scalability

A key engineering consideration for sequential refinement is that encoder inference must run over the full corpus before BiLSTM training can begin. Rather than treating this as a bottleneck, we exploit it as a decoupling opportunity: each encoder's representations are extracted and stored independently, and the BiLSTM trainer streams one encoder at a time while accumulating the ensemble prior as a running weighted sum. This means the BiLSTM can be retrained, ablated, or recalibrated at negligible additional cost — a property we used extensively during threshold search and loss configuration experiments.

All experiments use bf16 mixed precision throughout; random seed 42 is fixed for reproducibility. The complete system is containerised and submitted via TIRA [13].

10. Results

10.1. Main Results

Table 3 reports validation and official hidden test-set macro-F1 scores. Differences between the two splits are at most 0.005 across all three difficulties, indicating that the pipeline generalises reliably and that validation threshold tuning did not cause over-fitting.

10.2. Ablation Study

Table 4 decomposes the contribution of each pipeline stage on the validation set. The baseline uses a single best-performing encoder (DeBERTa) with its MLP classifier.

Table 3
Macro-F1 on the validation set and the official TIRA hidden test set.

Difficulty	Validation F1	Test F1
Easy	0.9931	0.9930
Medium	0.7270	0.7218
Hard	0.7550	0.7573

Table 4
Ablation: validation macro-F1 as pipeline components are added incrementally. Baseline uses a single DeBERTa MLP classifier.

Configuration	Medium F1	Hard F1
Single encoder (DeBERTa) MLP	0.674	0.735
+ Ensemble fusion	0.701	0.742
+ BiLSTM refinement	**0.727**	**0.755**

Ensemble fusion contributes +0.027 F1 on medium and +0.007 on hard by combining heterogeneous embedding spaces. BiLSTM refinement provides the larger gains (+0.026 on medium, +0.013 on hard), confirming that document-level transition structure is most valuable precisely where local pairwise evidence is weakest. The easy subset shows negligible change from either component ($<$0.001 F1), consistent with the view that high-contrast style differences are captured adequately by pairwise features alone.

10.3. Per-Difficulty Analysis

Easy. The system achieves 0.9930 test macro-F1, essentially at ceiling. Large inter-author style contrasts cause embeddings across boundaries to be naturally well-separated even before fine-tuning, and the BiLSTM adds negligible benefit. This suggests the easy subset is largely solved by current pretrained sentence representations.

Medium. At 0.7218 test macro-F1, medium is the most challenging result to achieve. Extreme positive weighting (w^+=25), a very low initial threshold (τ_0=0.08), R-Drop regularisation, and the consecutive-change penalty are all load-bearing components. The validation-to-test drop of 0.005 is within normal variance and does not indicate over-fitting.

Hard. The slight improvement from 0.7550 (validation) to 0.7573 (test) on hard suggests the BiLSTM's document-level modelling generalises particularly well to unseen documents with subtle style differences. This subset also shows the largest ablation impact from BiLSTM removal (0.024 F1), confirming that sequential context is essential when pairwise evidence is ambiguous.

11. Conclusion

We presented a four-stage system for sentence-level style change detection combining CoSENT-fine-tuned heterogeneous encoders, difficulty-adaptive MLP classifiers, Optuna-tuned ensemble fusion, and a BiLSTM that models document-level boundary transition structure. The system achieves near-ceiling performance on the easy subset and competitive results on the more demanding medium and hard splits, with stable generalisation to the official blind test set.

Several findings carry broader implications. The over-separation phenomenon during CoSENT training suggests that metric learning for short-text style discrimination benefits from early stopping against downstream task performance rather than the ranking loss itself. The consistent BiLSTM improvement confirms that document-level sequential context is a valuable signal at sentence granularity,

one that becomes more important as SCD datasets move to finer levels of analysis. Future work could explore end-to-end joint training of encoder and BiLSTM components, and investigate whether cross-difficulty transfer further improves medium-subset performance.

Acknowledgments

The authors thank the International Institute of Information Technology Hyderabad (IIIT-H) for computational infrastructure and research support. This work was supported through institutional start-up and SEED grants.

Declaration on Generative AI

During the preparation of this work the authors used Claude (Anthropic) in order to assist with LaTeX formatting and draft structuring. After using this tool the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. Bevendorff, M. Fröbe, A. Greiner-Petter, A. Jakoby, M. Mayerl, P. Nakov, H. Plutz, M. Potthast, B. Stein, M. N. Ta, Y. Wang, E. Zangerle, in: M. Hagen, M. Potthast, B. Stein, P. Schaer, E. Zangerle, A. Barrón-Cedeño, A. G. S. de Herrera, E. S. Salido, S. MacAvaney, J. M. Struß (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Seventeenth International Conference of the CLEF Association (CLEF 2026), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2026.

[2] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Multi-Author Writing Style Analysis Task at PAN 2026, in: A. Barrón-Cedeño, A. G. S. de Herrera, E. S. Salido, S. MacAvaney, J. M. Struß (Eds.), Working Notes of CLEF 2026 – Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, CEUR-WS.org, 2026.

[3] B. Bönninghoff, D. Hessler, A. Paeschke, R. M. Nickel, Deep hybrid long-short term memory network for authorship attribution, in: Proceedings of the Text2Story Workshop @ ECIR 2021, 2021.

[4] E. Stamatatos, et al., Overview of the PAN 2017 software and language forensics shared tasks, in: Working Notes of CLEF 2017, 2017.

[5] J. Su, CoSENT: Consistent sentence embedding via pair-wise cosine sentence loss, https://kexue.fm/archives/8847, 2022. Blog post.

[6] J. Weerasinghe, R. Greenstadt, Feature vector difference based neural network and logistic regression models for authorship verification, in: Working Notes of CLEF 2021, 2021.

[7] M. A. Hearst, TextTiling: Segmenting text into multi-paragraph subtopic passages, Computational Linguistics 23 (1997) 33–64.

[8] O. Koshorek, A. Cohen, N. Uzan, M. Rotman, J. Berant, Text segmentation as a supervised learning task, in: Proceedings of NAACL 2018, 2018, pp. 469–473. doi:10.18653/v1/N18-2075.

[9] A. Conneau, D. Kiela, H. Schwentz, L. Barrault, A. Bordes, Supervised learning of universal sentence representations from natural language inference data, in: Proceedings of EMNLP 2017, 2017, pp. 670–680. doi:10.18653/v1/D17-1070.

[10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of ICCV 2017, 2017, pp. 2980–2988. doi:10.1109/ICCV.2017.324.

[11] X. Liang, L. Wu, J. Li, Y. Wang, Y. Meng, X. Qiu, W. Zhou, G. Chen, X. Huang, R-Drop: Regularized dropout for neural networks, in: Advances in Neural Information Processing Systems (NeurIPS 2021), volume 34, 2021.

[12] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of KDD 2019, 2019, pp. 2623–2631. doi:10.1145/3292500.3330701.

[13] 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.