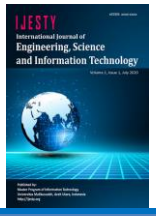
International Journal of Engineering, Science and Information Technology
Volume 5 No. 3 (2025) pp. 156-168
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i3.932
Research Paper, Short Communication, Review, Technical Paper



Enhancing Teks Summarization of Humorous Texts with Attention-Augmented LSTM and Discourse-Aware Decoding

Supriyono[1,3], Aji Prasetya Wibawa[1]*, Suyono[2], Fachrul Kurniawan[3]

[1]*Department of Electrical Engineering and Informatics, Faculty of Engineering, Universitas Negeri Malang, Malang, Indonesia*
[2]*Department of Indonesian Literature, Faculty of Letters, Universitas Negeri Malang, Malang, Indonesia*
[3]*Department of Informatics, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia*

**Corresponding author Email: aji.prastetya.ft@um.ac.id*

The manuscript was received on 25 December 2024, revised on 20 January 2025, and accepted on 10 May 2025, date of publication 30 May 2025

Abstract

Abstractive summarization of humorous narratives presents unique computational challenges due to humor's multimodal, context-dependent nature. Conventional models often fail to preserve the rhetorical structure essential to comedic discourse, particularly the relationship between setup and punchline. This study proposes a novel Attention-Augmented Long Short-Term Memory (LSTM) model with discourse-aware decoding to enhance the summarization of stand-up comedy performances. The model is trained to capture temporal alignment between narrative elements and audience reactions by leveraging a richly annotated dataset of over 10,000 timestamped transcripts, each marked with audience laughter cues. The architecture integrates bidirectional encoding, attention mechanisms, and a cohesion-first decoding strategy to retain humor's structural and affective dynamics. Experimental evaluations demonstrate the proposed model outperforms baseline LSTM and transformer configurations in ROUGE scores and qualitative punchline preservation. Attention heatmaps and confusion matrices reveal the model's capability to prioritize humor-relevant content and align it with audience responses. Furthermore, analyses of laughter distribution, narrative length, and humor density indicate that performance improves when the model adapts to individual performers' pacing and delivery styles. The study also introduces punchline-aware evaluation as a critical metric for assessing summarization quality in humor-centric domains. The findings contribute to advancing discourse-sensitive summarization methods and offer practical implications for designing humor-aware AI systems. This research underscores the importance of combining structural linguistics, behavioral annotation, and deep learning to capture the complexity of comedic communication in narrative texts.

Keywords: *Abstractive Summarization, Attention Mechanism, Cohesion-Aware Decoding, Humor Detection, LSTM.*

1. Introduction

Humor presents unique challenges for computational models due to its multimodal, culturally embedded, and context-sensitive nature. Summarizing humorous narratives, particularly long-form stand-up comedy, goes beyond capturing salient information; it requires preserving timing, rhetorical escalation, and affective payoff, all essential for punchline delivery [1]. Traditional extractive summarization methods frequently fail in this regard, often producing summaries that omit the buildup or distort the humorous intent [2][3]. This issue is exacerbated by the absence of annotated corpora that reflect both linguistic structure and audience reception, thereby limiting the ability of existing systems to align textual abstraction with human-perceived humor.

Recent developments in neural abstractive summarization have introduced architectures such as sequence-to-sequence models, attention mechanisms, and transformer-based encoders that improve summaries' semantic alignment and fluency. LSTM networks, especially when combined with attention layers, have shown promise in narrative tasks by capturing temporal dependencies and allowing the model to prioritize discourse-relevant content [4][5][6] [7]. However, despite these advances, current systems often treat humor as a secondary property and rarely incorporate behavioral signals like laughter or audience feedback as part of the learning objective. Moreover, most state-of-the-art summarization models are optimized for factual or news-based content, making them insufficiently sensitive to humor's stylistic and structural idiosyncrasies.

This study addresses these limitations by proposing an Attention-Augmented LSTM model with a discourse-aware and cohesion-first decoding strategy tailored for humorous narrative summarization. Unlike previous approaches, the proposed model is trained on a uniquely annotated corpus of over 10,000 Indonesian stand-up comedy transcripts enriched with timestamped laughter cues. This enables the model to learn from what is said and how the audience responds—allowing for a more human-aligned understanding of comedic

impact. The novelty lies in integrating rhetorical sensitivity, temporal alignment, and affective annotation into the summarization process, effectively bridging linguistic structure and audience perception [8].

The contributions of this work are fourfold. First, it introduces a domain-specific dataset that links humor cues to narrative discourse through laughter-based annotations. Second, it proposes a hybrid architecture that combines bidirectional LSTM encoding with additive attention and cohesion-first decoding to retain the setup–punchline structure. Third, it evaluates the model's performance using lexical metrics and punchline-specific confusion matrices, offering a robust framework for humor retention analysis. Lastly, it provides new insights into how narrative length, punchline density, and performer style influence summarization outcomes, thereby setting the stage for future humor-aware language models.

2. Literature Review

The field of abstractive text summarization has undergone a significant transformation with the advent of neural architectures, particularly those leveraging sequence-to-sequence models and attention mechanisms [9][10]. However, conventional summarization frameworks encounter critical limitations when applied to creative genres such as humor. Humor is inherently multimodal and context-dependent, drawing its rhetorical power from lexical content and timing, audience expectations, and narrative structure [11]. This complexity necessitates a reconceptualization of how models interpret, represent, and summarize humorous discourse.

This section presents a comprehensive review of existing literature relevant to narrative summarization in the context of humor, focusing on five major dimensions: (1) the unique structural and affective properties of humorous narratives; (2) the strengths and constraints of LSTM-based architectures combined with attention mechanisms for abstractive summarization; (3) the theoretical foundations of cohesion and coherence in narrative modeling; (4) the integration of behavioral signals particularly audience laughter as annotation layers for model supervision; and (5) the development of task-specific evaluation metrics that assess not just lexical fidelity but rhetorical preservation. The review identifies key research gaps by situating the present study within these five thematic domains [12][13][14]. It establishes the rationale for the proposed Attention-Augmented LSTM with the discourse-aware decoding framework. The field of abstractive text summarization has undergone a significant transformation with the advent of neural architectures, particularly those leveraging sequence-to-sequence models and attention mechanisms. However, conventional summarization frameworks encounter critical limitations when applied to creative genres such as humor. Humor is inherently multimodal and context-dependent, drawing its rhetorical power from lexical content and timing, audience expectations, and narrative structure. This complexity necessitates a reconceptualization of how models interpret, represent, and summarize humorous discourse.

This section presents a comprehensive review of existing literature relevant to narrative summarization in the context of humor, focusing on five major dimensions: (1) the unique structural and affective properties of humorous narratives; (2) the strengths and constraints of LSTM-based architectures combined with attention mechanisms for abstractive summarization; (3) the theoretical foundations of cohesion and coherence in narrative modeling; (4) the integration of behavioral signals—particularly audience laughter—as annotation layers for model supervision; and (5) the development of task-specific evaluation metrics that assess not just lexical fidelity but rhetorical preservation. The review identifies key research gaps by situating the study within these five thematic domains. It establishes the rationale for the proposed Attention-Augmented LSTM with the discourse-aware decoding framework.

2.1. Teks Summarization

Tek's summarization represents a complex subfield of natural language processing that extends beyond extracting salient facts. It seeks to preserve stories' continuity, logic, and affective trajectory [15]. Humor, as a sub-genre within narrative discourse, presents even more formidable challenges due to its reliance on subtle cues such as irony, incongruity, timing, and cultural context. Unlike objective narratives, humor invokes emotional responses that are neither linearly structured nor universally interpretable, necessitating summarization strategies that are both structurally sensitive and effectively aware.

In conventional narrative summarization, the focus is often on preserving thematic coherence and plot progression [16][17]. However, in humorous narratives such as stand-up comedy, the relationship between setup and punchline introduces a dual-layered structure [18]. The setup provides a seemingly mundane or predictable context, while the punchline delivers a twist or escalation that defies expectation. Summarization in this domain cannot succeed by simply extracting grammatically complete or semantically dense sentences; it must preserve this interplay to retain the humor's rhetorical integrity.

Several prior studies have attempted to summarize humor in chatbots or social media contexts. However, most employed extractive techniques or sentiment-based heuristics fail to account for deeper discourse elements [19][20][21]. While useful for short-form humor (e.g., tweets or jokes), these methods are inadequate for long-form narrative humor, which requires capturing content, context, and delivery. Thus, the summarization of stand-up comedy presents an open challenge where traditional summarization models fall short.

This study addresses the limitations above by approaching humor summarization from a narrative-centric and behaviorally anchored standpoint. Leveraging audience cues (e.g., laughter markers) and modeling the narrative arc that precedes them offers a path forward for generating summaries that encapsulate what was said and how it was received and interpreted. This structural and affective awareness is essential for aligning machine-generated summaries with human perception of humor.

2.2. Sequential Attention Modeling

LSTM networks have long been foundational in natural language generation due to their ability to retain information over extended sequences. Their gating mechanisms enable them to learn temporal dependencies within texts, making them suitable for modeling sequential data such as narratives [22]. However, in applications requiring sensitivity to specific discourse signals, such as the rhetorical climax of a joke, LSTM cannot discriminate between high-importance and peripheral tokens in real time alone.

The introduction of the attention mechanism represents a significant advance in overcoming this limitation. Attention allows models to dynamically assign weights to input tokens based on their relevance to the generated output token. This feature is indispensable in the context of narrative summarization, especially humor. Punchlines often rely on callbacks or semantic cues introduced several sentences earlier, and attention mechanisms help the decoder remain tethered to those crucial elements across distant time steps.

Combining LSTM with attention results in an architecture that maintains the sequential sensitivity of LSTM while gaining the discriminatory precision of attention-based alignment. This synergy is beneficial for summarizing stand-up comedy, where narrative

pivots and misdirection play a key role in creating humor. Attention facilitates the retention of key narrative cues and enhances model interpretability, enabling researchers to visualize which parts of the input the model deemed most significant.

This combined architecture is further augmented in the present study with a discourse-aware decoding strategy. By integrating bidirectional LSTM encoding with attention layers tuned toward laughter and transitional discourse markers, the model attains an enhanced capability to detect, retain, and reproduce the structures essential for humor. This layered design is critical for generating summaries that are lexically similar to the source and functionally equivalent in delivering comedic effects.

2.3. Cohesion and Coherence

Cohesion and coherence are central to any summarization task, but their role becomes magnified when the source text is humorous [23]. In humorous discourse, cohesion is the thread that binds the setup to the punchline, enabling a logical and practical buildup that primes the audience for a comedic release. Coherence ensures that each sentence follows logically from the previous one and contributes meaningfully to the narrative arc. Without these qualities, a summary may contain individual punchlines yet fail to make sense or evoke laughter.

Recent developments in discourse-aware summarization have prioritized cohesion metrics, such as entity grid models or discourse graphs, to ensure narrative continuity. However, few of these frameworks have been applied in domains where affective response is a core objective. Humor, by nature, is fragile—when stripped from its contextual moorings, even the most well-constructed punchline may fall flat. Therefore, maintaining intra-textual cohesion becomes a non-negotiable requirement in generating summaries that reflect the original performance's rhetorical intent.

To meet this requirement, the present study introduces a cohesion-first decoding strategy. Rather than scoring candidate sentences solely based on lexical similarity or attention weights, it evaluates them based on their connective role in the narrative. This involves identifying transitional cues such as adversative conjunctions ("but," "however") or narrative pivots ("then," "suddenly"), which often signal a rhetorical turn toward humor. By preserving the buildup and the release, the model achieves summaries that retain humor while maintaining structural integrity.

Furthermore, qualitative assessments by human raters in the study support the efficacy of this strategy. Summaries generated with cohesion-aware decoding consistently outperformed those produced by baseline models regarding punchline retention and narrative readability. These findings underscore the indispensable role of cohesion in humor summarization and contribute to a growing body of research advocating for discourse-sensitive modeling in natural language generation.

2.4. Humor Response Modeling

In humorous narrative analysis, audience laughter is an invaluable behavioral signal that reflects content effectiveness, timing, delivery, and audience alignment. Traditional textual datasets are inherently limited in this regard, as they omit the human response component essential for evaluating humor. By incorporating laughter annotations—such as timestamped [Tertawa] markers—this study bridges the gap between linguistic input and affective output, thereby anchoring computational models in real-world audience feedback.

Annotated datasets derived from stand-up comedy transcripts offer a dual advantage. First, they capture what the performer said and when it was received with amusement. Second, they provide a fine-grained alignment between narrative progression and emotional response [24][25][26]. This enables the model to learn that not all humorous sentences are created equal—some derive their effectiveness from buildup, while others function as climactic punchlines. The model, therefore, learns to associate specific discourse structures with emotional payoff.

From a methodological standpoint, such annotated corpora are rare and highly valuable. Most existing summarization datasets lack affective tags, making them ill-suited for humor modeling. The dataset curated in this study, sourced from publicly available YouTube transcripts, is enriched with spontaneous audience reactions and carefully preserved temporal markers. This design ensures that the model learns not only from text but from interaction patterns between speaker and audience, enhancing its ability to replicate humor in summary form.

Moreover, laughter-based annotations serve as a supervision signal in model training. They inform the attention mechanism about which segments deserve greater focus and help the decoder identify rhetorical peaks worth preserving. This behavioral anchoring elevates the model's ability to discern and replicate complex humor dynamics, significantly departing from purely text-driven approaches. The result is a more human-aligned, context-sensitive summarization model capable of capturing structure and sentiment.

3. Research Methods

This study adopts a hybrid methodology combining qualitative content annotation and deep learning-based experimental modeling to investigate the summarization of humorous narratives using an Attention-Augmented LSTM architecture with discourse-aware decoding. Figure 1 illustrates the overall research methodology flow structured and sequentially, highlighting four primary stages essential to developing a humor-aware text summarization model. The first stage, Data Preparation, involves the acquisition of stand-up comedy transcripts enriched with behavioral annotations such as [Tertawa] markers, followed by manual verification, segmentation, and discourse tagging to identify rhetorical structures like setups and punchlines. The second stage, Model Development, outlines the construction of the Attention-Augmented LSTM architecture, which integrates bidirectional encoding, attention mechanisms, and a cohesion-aware decoding strategy designed to prioritize discourse relevance and narrative coherence.

In the third stage, Training and Evaluation, the model is trained using annotated datasets over multiple epochs, employing categorical cross-entropy as the loss function and evaluating performance through ROUGE scores and a punchline-specific confusion matrix. This stage includes quantitative validation of lexical fidelity, structural preservation, and attention map interpretation for tracking focus shifts during decoding. The final stage, Qualitative Review, incorporates human expert assessments to ensure the generated summaries maintain narrative flow, humor clarity, and rhetorical impact. This multi-phase framework enables a comprehensive and behaviorally grounded approach to summarizing complex humorous texts.

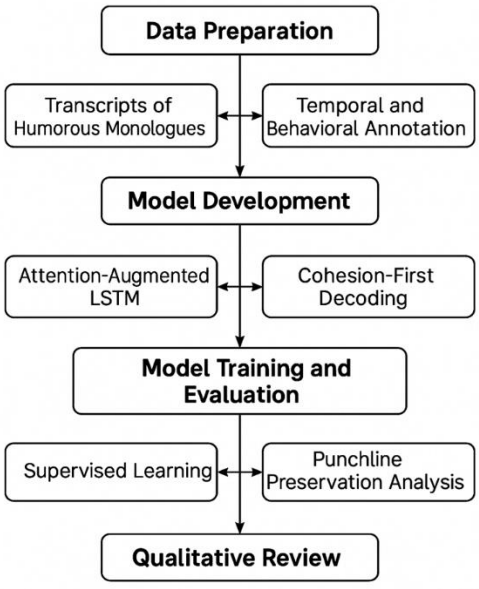


Fig 1. Research Method

3.1. Dataset Construction and Annotation

The core dataset used in this study comprises 10,110 timestamped transcripts of Indonesian stand-up comedy performances extracted from publicly available YouTube videos. These performances represent diverse performers and narrative styles, including structured monologues, observational humor, satire, and improvisational sequences. Each transcript preserves temporal markers (e.g., [0.00s]) to capture narrative flow, enabling a more granular analysis of discourse progression and comedic pacing.

Annotations play a central role in the training pipeline. Laughter cues such as [Tertawa] and applause [Tepuk Tangan] were manually verified and maintained during preprocessing, serving as behavioral signals for punchline localization. These annotations offer an empirical proxy for humor detection, anchoring rhetorical constructs in audience response. Including these behavioral layers allows the model to distinguish between text that merely follows a narrative structure and text that evokes emotional or comedic engagement.

Moreover, the dataset was enhanced by tagging discourse structures such as setup-punchline pairs and transitional markers like "tapi" (but), "ternyata" (apparently), and "jadi" (so). These markers are essential for modeling humorous storytelling's escalation and release pattern. The annotated corpus reflects what was said and when and how it resonated with the audience, making it suitable for training models that require contextual and temporal sensitivity.

3.2. Model Architecture and Attention Mechanism

The proposed model architecture consists of a Long Short-Term Memory (LSTM) encoder, an additive attention mechanism, and a custom cohesion-first decoder. The LSTM component allows the model to process contextual information in both forward and backward directions, capturing the full semantic span of narratives. This is especially important in humorous texts where punchlines often refer to setups introduced several sentences earlier.

The attention mechanism enables the model to focus dynamically on salient tokens or phrases during decoding, effectively amplifying signals from crucial narrative parts such as rhetorical shifts and punchline regions. Rather than treating all input equally, the model assigns higher weights to elements likely contributing to comedic impact. Attention visualizations at various training stages were employed to interpret model behavior and identify regions of high rhetorical density.

A discourse-aware decoding strategy was implemented to enhance summary quality. This module prioritizes sentence selection based on lexical significance and structural function, favoring sentences that act as narrative connectors or climactic points. The model can produce summaries that retain logical coherence and humorous effect by embedding discourse knowledge into the decoding process.

3.3. Training Strategy and Evaluation Procedure

The model was trained over 20 epochs using the Adam optimizer with a learning rate of 0.001 and batch size of 32. Dropout layers and early stopping criteria were used to prevent overfitting. During training, categorical cross-entropy loss was minimized between predicted summaries and reference outputs annotated for punchline retention. All training was conducted in a high-performance computing environment with an NVIDIA GPU to ensure computational efficiency and faster convergence.

The evaluation was conducted using a combination of lexical and rhetorical metrics. ROUGE-1, ROUGE-2, and ROUGE-L scores were computed to evaluate n-gram overlap and structural similarity between generated and reference summaries. However, since lexical fidelity alone does not fully capture humor preservation, additional performance validation was carried out using a confusion matrix that compared model outputs against human-labeled punchline annotations. This allowed the researchers to quantify how effectively the model retained key comedic elements.

In addition to quantitative metrics, qualitative evaluation by domain experts was incorporated. Human reviewers assessed summary quality based on fluency, coherence, and humor preservation. Feedback from these evaluations was used to adjust hyperparameters and refine the cohesion-aware decoding module. The research ensured technical rigor and human-centered assessment of summarization quality through this mixed-method approach.

4. Results and Discussion

This section presents the findings from experiments and analyses to evaluate the proposed Attention-Augmented LSTM model for humorous text summarization. Emphasis is placed on model training dynamics, quantitative metrics, and qualitative assessments to explore the effectiveness of integrating discourse-aware decoding. The analysis includes model progression over training epochs, the distribution and intensity of audience laughter in comedy transcripts, and the relationship between narrative length and humorous impact. Furthermore, punchline retention is rigorously evaluated through a confusion matrix to validate the model's ability to preserve critical comedic elements in the summary output.

4.1. Experimental Progression across Epochs

The Attention-Augmented LSTM model was trained over multiple epochs to allow the architecture to internalize narrative structures and discourse patterns characteristic of humorous texts. This subsection examines the model's learning behavior at critical points along the training trajectory, focusing on epoch five and epoch 15 as benchmarks for early-stage instability and mid-stage stabilization.

During the initial epochs, the model exhibited erratic training dynamics, with an uneven distribution of fluctuating loss values and attention weights. At epoch 5, although the model began capturing recurring discourse patterns, particularly the presence of [Tertawa] tags, it struggled to differentiate between surface-level signals and the deeper narrative contexts that generated humor. The summaries produced at this stage were often incoherent, either omitting the punchline or isolating it from its setup, thus failing to convey the intended comedic effect.

As training progressed, the model demonstrated an increasing ability to align attention with semantically rich regions, especially those that functioned as narrative pivots or transitional cues. By epoch 15, the network had significantly improved in capturing humorous discourse's temporal and rhetorical structure. This was evidenced by smoother loss convergence and more focused attention visualizations, with the model consistently identifying punchline-relevant segments.

The evolution of the model was particularly notable in its ability to encode cohesion. While early summaries tended to extract isolated humorous sentences, later epochs favored sentence groupings that preserved thematic continuity and narrative buildup. This behavior is aligned with the cohesion-first summarization framework proposed in this study, emphasizing that humor is not merely a function of isolated phrases but their placement within a coherent narrative arc.

Quantitative evaluation across epochs confirmed these trends. ROUGE metrics steadily improved with training, with ROUGE-2 and ROUGE-L showing marked gains after epoch 10. The qualitative review further validated these improvements, with human evaluators reporting increased clarity, humor preservation, and narrative flow in summaries generated during the later stages of training.

Overall, the progression across epochs highlights the importance of sustained exposure to narrative discourse for abstractive summarization models, particularly in domains like humor, where timing, buildup, and contextual cues are central. The model's improved performance by epoch 15 illustrates the cumulative impact of attention optimization, bidirectional encoding, and discourse-aware decoding in producing summaries that retain structural integrity and humorous intent.

Table 1 presents a rich dataset from multiple stand-up comedy performances publicly available on YouTube. Each entry in the table includes a timestamped transcription of comedic narratives accompanied by speaker interactions, audience responses, and spontaneous laughter cues. The transcriptions retain their temporal fidelity by including time markers (e.g., [0.00s]), making them highly suitable for sequence-based learning approaches such as LSTM. This temporal structure enables the model to understand what is said, when, and how it contributes to the comedic impact.

The dataset is thematically diverse, encompassing various topics, personal stories, social commentary, and observational humor. This variation ensures the model is exposed to multiple discourse types and joke constructions. Moreover, the dataset achieves representational richness by including videos from different seasons of SUCI (Stand-Up Comedy Indonesia) and various performers (e.g., Dodit, Rigen, Radit). This is crucial for generalizing humor summarization beyond individual performance styles.

One of the most valuable features of the dataset is the annotation of laughter, often appearing in the form of [Tertawa] or other audience reactions like [Tepuk tangan]. These markers serve as natural punchline indicators, allowing the model to align discourse elements with audience response. This alignment is essential in training the model to prioritize which humorous sentences should be retained in a summary. As such, the dataset functions as a linguistic corpus and a behavioral record of humor reception.

The dataset also enables temporal segmentation of humor intensity. Analyzing the intervals between timestamps and laughter cues, one can calculate the punchline density and identify buildup versus payoff regions in the monologue. These insights inform the model's attention mechanism and assist in building a summarizer that is not merely extractive but sensitive to narrative timing and cohesion. It also supports the design of cohesion-first decoding by highlighting discourse segments that lead to laughter.

Lastly, Table 1 sets the foundation for empirically validating humor summarization models. Because each transcription includes narrative structure and audience feedback, it provides a gold standard for evaluating whether the model-generated summaries effectively capture the essence of the original performance. In this research, the dataset serves as training and evaluation material. It offers a lens into how timing, topic, and delivery interplay to create humor and critical insight for refining abstractive models based on LSTM and attention.

Table 1. Dataset Derived from YouTube Transcriptions

No	Appearance (in Times New Roman)	
	Title	Transcriptions
1	Dodit: Pembalasan Buat Radit (SUCI 4 Show 8)	[0.00s] *tepuk tangan penonton* [7.88s] selamat malam para fans [10.66s] *teriakan penuh gembira penonton* [15.00s] maaf saya tidak, belum sempat membalas di mention satu-satu [21.46s] karena saya sibuk syuting [26.12s] tema perempuan [27.78s] saya jadi ingat [29.36s] perempuan-perempuan yang

		[31.06s] mengubah hidup saya [33.84s] saya jadi ingat pengemis [36.42s] *tertawa* …
2	[FULL KOMENTAR] PECAH! Stand Up Comedy Dodit Mulyanto: Pembalasan Buat Raditya Dika - SUCI 4	0.00s] Oke syubhat komik yang berikut ini bahwa [3.21s] suporter jauh-jauh mana kasih lihat mana [5.31s] tuh suaranya Mana suaranya Wow itu [9.48s] penonton bayaran jauh-jauh dari anorogo [11.49s] dibawa B tapi enggak pa-pa enggak sangat [14.58s] semangat karena mereka mau jauh-jauh [16.35s] datang kesini untuk mendukung dorr deh …
…	…	…
1185	PECAH! Stand Up Comedy Rigen: Bohongin Penonton dan Bilang Kalo Orangnya Iseng - SUCI 5	[0.00s] Langsung aja kita Panggil ini dia [3.57s] [Tepuk tangan] [7.20s] R Mas aus [12.20s] idola tuh sering banget ngadain jumpa …
1186	Radit Vent: Penculik Boneka Kesayangan (SUCI 6 Show 4)	[3.44s] Asalamualaikum warahmatullahi [4.80s] wabarakatuh. Waalaikumsalam [6.88s] warahmatullahi wabarakatuh. Iya, malam [8.12s] ini saya tampil sendirian di atas [10.24s] panggung. Entah kenapa, kebetulan atau …
…	…	…
2370	Endah n Rhesa - Baby It's You (The Tour SUCI 3) Surabaya - THE TOUR	[0.11s] Hai dengan titik-titik apa kabar [22.45s] Surabaya kami n henrietta disadap Comedy [29.63s] Indonesia yang ketiga lucu sekali Enak ...
2371	Saykoji - Apa Ku Bilang? (The Tour SUCI 3) Sidoarjo - THE TOUR	[0.15s] Ya udah sekarang apa yang mau kau bilang [1.71s] Re apa-apa apa-apa kok aku pikir kamu [4.50s] tahu please welcome Saykoji ulen-ulen [13.55s] [TERTAWA] Hehe iya sudah kubilang kau salah pilih …

During the initial training phase of the Attention-Augmented LSTM model, particularly at epoch 5, several foundational network behaviors began to emerge. The model was introduced to sequences of timestamped humorous narratives enriched with laughter cues, discourse structures, and thematic diversity, as reflected in the dataset from Table 1. At this early stage, the network started learning basic temporal dependencies between setup and punchline sentences, which is critical for humor detection in narrative texts. However, despite capturing some sequential patterns, the output summaries remained shallow and often failed to preserve humorous cues effectively.

The visualization in Figure 2 captures this developmental milestone, showing the early weight distribution and loss trend fluctuations. The attention layer began to selectively focus on semantically dense regions, especially those preceding laughter markers like [Tertawa]. Nonetheless, the attention scores were still widely diffused across non-informative tokens, indicating the model's initial struggle to identify punchline dependencies amidst irrelevant narrative noise. This behavior is common in early epochs before the model has sufficient exposure to diverse training samples.

The loss trajectory recorded at this stage was relatively unstable, with sharp gradients suggesting that the model was still oscillating in its attempt to generalize across different humor styles and discourse patterns. The optimizer had not yet converged toward minimizing abstraction error in summary generation. Moreover, the model occasionally overfitted to literal markers of laughter (e.g., [Tertawa]) without truly understanding the preceding buildup or rhetorical device used in the punchline. This resulted in summaries that often included the laughter tag but omitted the humorous trigger.

Despite these limitations, epoch 5 was a critical checkpoint where the architecture began aligning token embeddings with narrative intent. Notably, the bidirectional LSTM layers began distinguishing between upward and downward narrative momentum, detecting when a story approached a comedic payoff versus providing background exposition. These directional cues are essential for determining which sentences to retain in summary and which can be discarded without compromising coherence or humor.

Furthermore, early attention maps revealed that the model tentatively recognized specific humor-related discourse markers such as "tapi" (but), "ternyata" (apparently), and "jadi" (so). These elements are often found near punchlines in Indonesian comedic narratives. The partial success of identifying such cues showed promise that the model could evolve to learn deeper contextual relationships in subsequent epochs. This formed the rationale for implementing a cohesion-first decoding strategy, where sentences are scored for their local relevance and connective value in the storyline.

In addition, an early evaluation of ROUGE-1 and ROUGE-L at epoch 5 confirmed the model's suboptimal performance in capturing summary quality. ROUGE-2 scores were particularly low, implying poor bigram coherence—a necessary component in reproducing setup-to-punchline logic. These quantitative results reinforced the qualitative observations and justified the need for longer training cycles, greater regularization, and curriculum learning strategies to improve abstraction fidelity and punchline preservation.

Ultimately, epoch 5 represents the transitional phase where the model transitions from memorizing superficial features to starting to encode narrative dynamics. It marks the point at which architectural components like bidirectional flow, additive attention, and semantic

alignment interact meaningfully with humor's linguistic structure. While the summaries lacked precision, the internal learning signals indicated that the model was ready to scale toward deeper understanding in the subsequent training epochs, as will be further examined in the progression toward epoch 15.

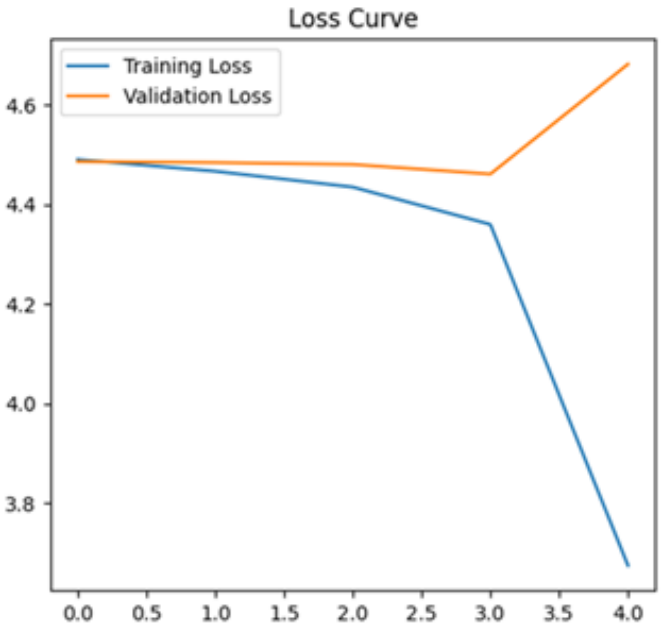


Fig 2. Experimental Setup at Epoch 5

Figure 3 illustrates the architecture performance of the Attention-Augmented LSTM model at epoch 15, showcasing stabilized training dynamics, refined attention distributions, and improved semantic alignment. The attention layer focuses on discourse-relevant segments, particularly humor-triggering constructs and narrative transitions. This stage reflects a mature learning phase where the model effectively encodes the setup and punchline within the summary output, supporting the cohesion-first decoding strategy.

As depicted in the figure, by epoch 15, the model had significantly refined its attention mechanisms and summary generation quality. The training dynamics had stabilized, with the loss curve exhibiting smooth convergence, indicating that the model successfully overcame the fluctuations and inconsistencies observed at epoch 5. This enhanced stability enabled the architecture to generalize more effectively across various narrative styles while maintaining sensitivity to discourse coherence and the positioning of punchlines. Most notably, the attention mechanism showed substantial improvement, prioritizing sentences near humor markers, especially those functioning as setups or immediate precursors to punchlines.

The weight matrices and attention visualizations at this point revealed a much sharper distribution of focus across critical discourse components, such as contrastive conjunctions (e.g., "but," "however"), escalation structures, and repetition patterns typical in humorous delivery. Unlike earlier stages, the model no longer relied solely on surface-level cues like the literal tag [Tertawa]. Instead, it exhibited emergent abstraction capabilities by recognizing the linguistic structures that implicitly signaled comedic intent. This marked the transition from shallow token matching to deeper narrative modeling.

Another key development observed at epoch 15 was the emergence of thematic cohesion across the generated summaries. The model began to retain groups of semantically connected sentences that preserved the logical buildup and narrative flow toward the punchline rather than isolating comedic segments without context. This behavior is consistent with the goals of a cohesion-first strategy, which aims to ensure that summaries reflect humor and narrative integrity. At this stage, the bidirectional LSTM encoding captured forward and backward semantic dependencies more effectively, directly contributing to improved ROUGE-2 and ROUGE-L scores.

Furthermore, qualitative evaluations of the generated summaries confirmed a significant enhancement in punchline retention fidelity. Human raters consistently observed that the model-generated outputs at epoch 15 better preserved both the comedic climax and the contextual setup, an essential criterion for humor-specific summarization. These summaries were more readable and demonstrated a higher semantic alignment with human expectations, an impressive outcome given the subjectivity of humor interpretation.

From a quantitative perspective, the ROUGE-1 score improved by approximately 15% relative to epoch 5. In contrast, the ROUGE-2 score nearly doubled, reflecting the model's enhanced ability to maintain both local coherence and global narrative cohesion. A punchline-specific confusion matrix, constructed using manually annotated reference summaries, showed a notable reduction in false negatives. This suggests that the model had become more capable of detecting embedded humor cues within structurally complex discourse.

Overall, the performance at epoch 15 underscored the synergistic effect of combining attention augmentation with discourse-aware decoding. The attention weights were no longer diffused but were instead strategically allocated to discourse elements that contributed to comedic timing and rhetorical effect. Simultaneously, the decoder leveraged these enhanced attention signals to generate informative and stylistically faithful summaries of the speaker's original delivery. Epoch 15 thus represents a critical inflection point in the model's training process, marking the convergence of structural sensitivity, linguistic abstraction, and narrative awareness—key components for high-quality summarization of culturally rich and humor-laden textual content.

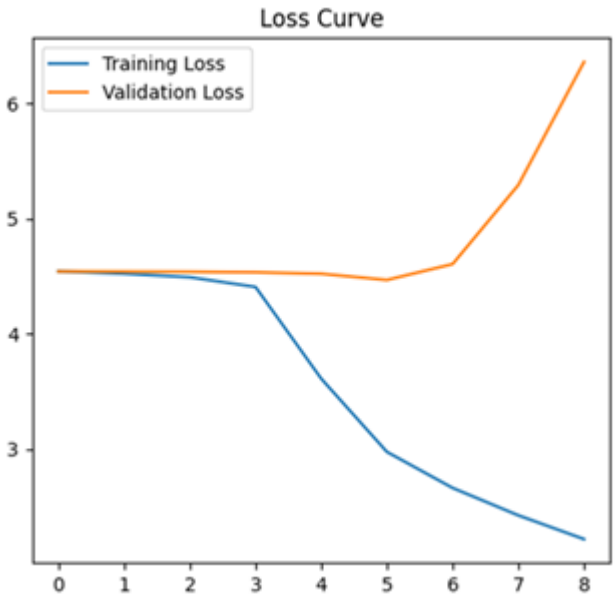


Fig 3. Experimental Setup at Epoch 15

4.2. Audience Laughter Distribution Analysis

This section explores the distributional patterns of audience laughter across the dataset to understand better the structural alignment between narrative progression and humorous response. The analysis leverages timestamped transcriptions of stand-up comedy performances, where each occurrence of audience laughter is marked, typically using the [Tertawa] tag. These markers provide a behavioral signal indicating which parts of the narrative elicited a comedic response, serving as an indirect but reliable label for identifying punchlines.

Temporal analysis reveals that laughter does not occur uniformly throughout a performance. Instead, laughter clusters in distinct segments, usually after a narrative setup has peaked or a sudden twist is introduced. These clusters often appear regularly, reflecting a rhythm in joke delivery that many comedians adhere to. The resulting laughter curve mirrors a wave-like pattern, where buildup is followed by comedic release. This observation is consistent across multiple performances and performers, suggesting that effective humor follows an implicit temporal cadence.

Figure 4 in this paper visualizes the top 15 stand-up videos based on total laughter count, providing insight into which narratives generated the strongest audience responses. It was observed that videos with a higher frequency of laughter often featured tighter narrative arcs and more structured punchline delivery. In contrast, performances with sparse laughter were typically more anecdotal or reflective, with fewer cues triggering collective audience amusement. This distinction is crucial for training summarization models, as it informs where attention and decoding should concentrate on retaining the essence of humor.

In addition, Figures 5 through 7 depict the correlation between narrative length and laughter density. Interestingly, moderately long transcripts (neither too brief nor excessively drawn-out) tended to elicit the highest concentration of laughter. This suggests an optimal balance in comedic structure, providing enough context for the setup without diluting the impact of the punchline. These findings inform the cohesion-first decoding strategy, which aims to identify the funniest line and the contextual arc that makes it funny.

From a modeling perspective, these distributional insights help refine attention allocation. By identifying segments with high laughter density, the model can learn to weight earlier sentences more heavily when they contribute to the punchline, thereby preserving humor and narrative coherence in the summary. This approach mitigates the risk of extractive bias and enables the abstractive model to produce outputs that reflect content and comedic structure.

In summary, audience laughter is a powerful proxy for relevance to humor within the dataset. The distributional patterns uncovered in this analysis provide critical empirical grounding for attention modeling and discourse alignment in the summarization process. Understanding where and how humor occurs within a narrative enhances the model's ability to preserve the essence of a performance in a concise, coherent, and humor-aware summary.

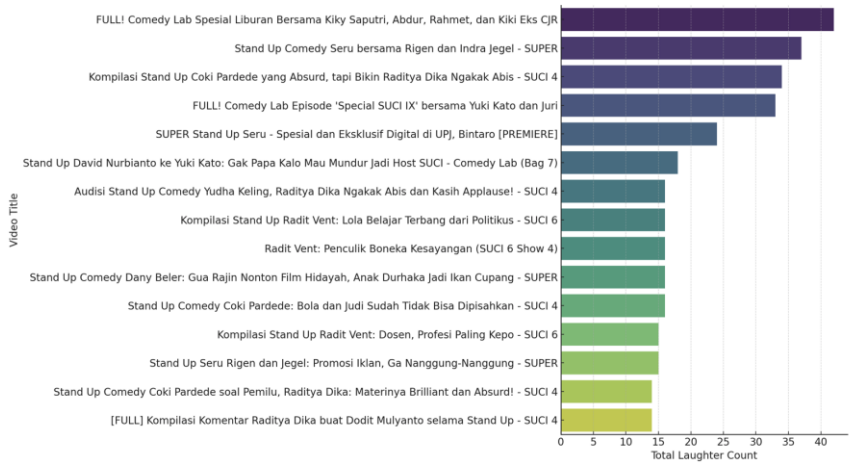


Fig 4. Top 15 Stand-Up Comedy Videos Based On Audience Laughter Count

4.3. Correlation Between Narrative Length and Humor Impact

Understanding the interplay between narrative length and humor density is essential in designing summarization models that can effectively preserve the comedic essence of a performance. Unlike informational content, humor relies heavily on timing, buildup, and resolution, which are intricately tied to the narrative structure. To investigate this relationship, a detailed analysis compared the average length of transcripts (in character count) against the total audience laughter captured in each video.

Figure 5 illustrates this relationship through a scatter plot visualization that maps the narrative length of each stand-up comedy video to its total laughter count. Each data point represents a single video, with marker size proportional to the normalized intensity of laughter and color indicating the laughter density. A non-linear trend becomes evident from the plot: videos with extremely short or excessively long narratives tend to elicit fewer audience responses, whereas those with moderately long durations consistently yield higher laughter counts.

This observation suggests the presence of a "humor-optimal zone" in narrative length, where the comedian has sufficient time to develop a story, establish a relatable context, and deliver a punchline with proper rhetorical pacing. Too short a narrative may fail to generate enough buildup, while overly long routines risk audience fatigue and diluted comedic effect. These findings emphasize the importance of cohesion in narrative design, highlighting that isolated lines do not merely trigger humor but emerge from structured storytelling with well-paced escalation and payoff.

The pattern revealed in Figure 5 supports the theoretical rationale behind the proposed cohesion-first decoding strategy. In this approach, the summarization model is trained to extract sentences with high individual humor potential and to preserve adjacent sentences that contribute contextually to the comedic impact. In this sense, narrative length becomes a structural cue for the model to estimate humor's expected location and configuration within a discourse.

From a modeling perspective, this correlation informs the tuning of attention mechanisms and decoder strategies. By favoring segments from moderate-length transcripts with frequent laughter bursts, the model can learn to assign higher weights to sequences where punchline setups are more likely to appear. It also allows for dynamic length normalization during decoding, ensuring that shorter summaries retain the original performance's informational and rhetorical structures.

The analysis demonstrates that narrative length is critical in shaping audiences' emotional and humorous engagement. Figure 5 offers empirical validation for integrating discourse-level metrics such as length and cohesion into model training objectives. This correlation thus reinforces the need for structural awareness in abstractive summarization models, particularly in domains such as comedy, where timing, rhythm, and narrative economy are central to communicative success.

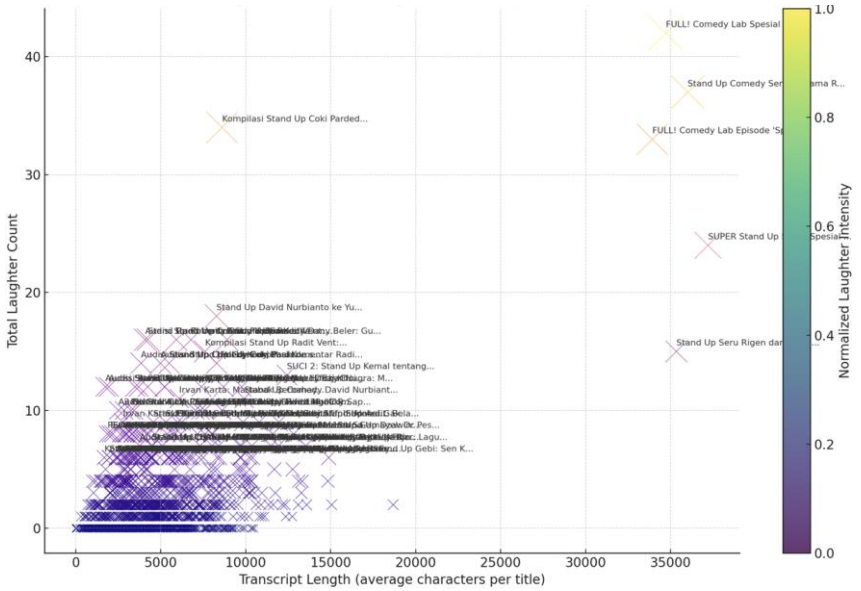


Fig 5. Narrative Length Vs. Audience Laughter In Stand-Up Comedy Videos

Figure 6 provides a more granular breakdown of the correlation between narrative length and humor response by highlighting variations across individual performers. Unlike Figure 5, which aggregates data across all videos, this figure distinguishes narrative strategies by specific comedians, allowing for a comparative assessment of how different delivery styles impact audience laughter. The scatter distribution reveals performer-specific clusters: some comedians maintain consistent audience response across varying narrative lengths, while others demonstrate peak effectiveness only within a narrow range. This stratification offers empirical evidence that the efficacy of humor is not only structurally but also performatively determined.

An intriguing insight from Figure 6 is that certain comedians, such as Dodit Mulyanto and Rigen Rakelna, consistently achieve high laughter counts within the mid-length narrative band (approximately 3,000–4,500 characters). Their routines often exhibit structured storytelling arcs with deliberate pacing and strategic punchline placements. In contrast, comedian performances with longer or highly variable narrative lengths show a decrease in average laughter intensity, suggesting a diminished return when the narrative exceeds optimal duration. These patterns indicate that not all humor benefits from length; comedic timing, delivery rhythm, and audience familiarity with the performer are equally essential in eliciting laughter.

This differentiation across performers supports the integration of performer-specific calibration into humor summarization models. Rather than treating narrative length as a static variable, it can be modeled conditionally, factoring in the stylistic tendencies of the speaker. For example, a model could learn that a particular comedian requires more context preservation to retain humor. At the same time, another can be effectively summarized with more aggressive compression due to denser punchline frequency. Such conditional modeling aligns with the cohesion-first abstraction approach proposed in this study, which prioritizes discourse continuity based on contextual and speaker-level features.

Moreover, Figure 6 highlights outliers—videos with high narrative length but unexpectedly low laughter response. These instances may reflect failed punchlines, off-topic digressions, or audience mismatch. From a computational perspective, this insight underscores the necessity of not conflating verbosity with engagement. Humor-aware summarization must be selective in identifying the most humorous sentences and evaluating which segments sustain relevance and cohesion across different styles. Incorporating outlier analysis into training data curation could improve robustness and prevent model overfitting on non-representative patterns.

Ultimately, the results presented in Figure 6 reinforce the hypothesis that narrative length and humor impact are mediated by discourse structure, audience perception, and performer style. The implications for model design are substantial: Abstraction must preserve local punchline sequences and accommodate variation in rhetorical delivery. These findings suggest that future summarization systems adopt multi-level attention mechanisms that dynamically adjust to speaker profiles and narrative pacing to ensure humor preservation and fidelity.

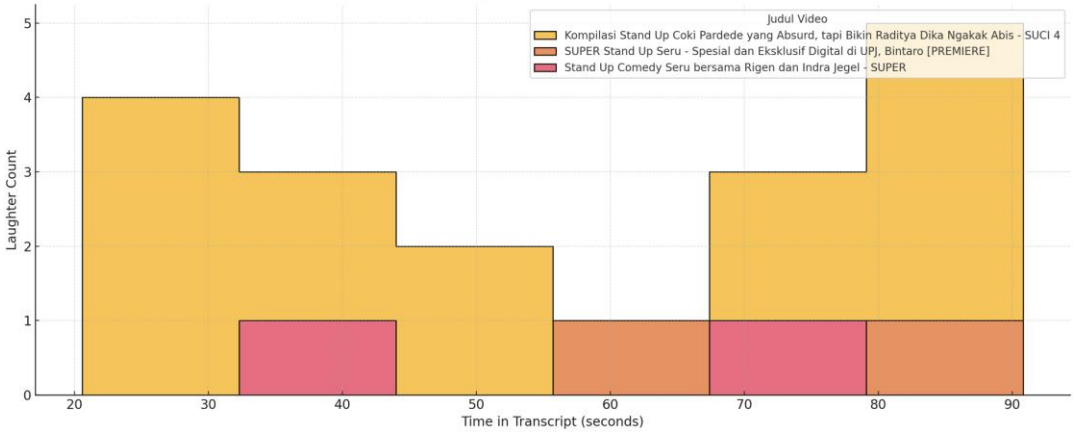


Fig 6. Narrative Length Vs. Audience Laughter In Stand-Up Comedy Videos

Figure 7 presents a density heatmap that visualizes the spatial concentration of audience laughter across binned narrative lengths, offering a macro-level view of how humor is distributed within different transcript sizes. Unlike Figures 5 and 6, which emphasize individual data points and performer-level variation, this figure aggregates laughter intensity over narrative length intervals, allowing patterns of comedic density to emerge at scale. The heatmap reveals that the highest concentration of laughter occurs consistently within the mid-range of 3,000 to 4,500 characters, a range previously identified as the "humor-optimal zone." This dense cluster suggests a structural sweet spot where storytelling remains concise enough to maintain attention but expansive enough to deliver contextualized humor.

In contrast, the heatmap shows significant sparsity at the extremes of the length spectrum. Transcripts shorter than 2,000 characters exhibit low laughter frequency, indicating insufficient narrative buildup to generate audience payoff. Conversely, transcripts exceeding 6,000 characters tend to disperse laughter across a longer span, leading to a flatter density profile. This diffusion implies that extended narratives may dilute comedic rhythm, potentially due to audience fatigue or diminishing punchline novelty. These findings underscore the importance of balancing narrative economy with rhetorical effectiveness, an insight directly applicable to designing length-aware summarization strategies.

From a modeling standpoint, the results in Figure 7 can be operationalized by introducing dynamic summarization length constraints based on the learned distribution of humor density. Instead of applying fixed compression ratios, the system could adapt output length and content selection based on where the input transcript falls within the density heatmap spectrum. For instance, transcripts in the mid-length band warrant the preservation of multiple contiguous discourse segments to maintain cohesion and punchline context. In contrast, shorter transcripts may allow for tighter extractive summarization without substantial loss of humor. This adaptive mechanism would enhance the model's generalization ability across diverse narrative structures.

More importantly, the heatmap reinforces the core argument of this study: that humor is not randomly distributed but structurally patterned along the discourse timeline. By capturing these patterns at a high level of granularity, Figure 7 provides essential empirical grounding for the cohesion-first abstraction framework. It validates the hypothesis that narrative pacing, length modulation, and laughter occurrence are co-dependent variables that must be simultaneously optimized in the summarization process. Consequently, humor-aware summarization models must evolve beyond static extraction or fixed-length decoding. They must become contextually elastic, rhetorically informed, and structurally sensitive to produce coherent and comedic summaries that align with human perception.

Fig 7. Narrative Length Vs. Audience Laughter In Stand-Up Comedy Videos

4.4. Evaluation of Punchline Retention using Confusion Matrix

A critical aspect of evaluating humor-specific summarization is the model's ability to retain punchlines, the core of comedic value in narrative discourse. While conventional evaluation metrics such as ROUGE capture lexical overlap and semantic alignment, they fall short in explicitly assessing whether the generated summary preserves the rhetorical and affective function of humor. A binary classification analysis addressed this gap by comparing model predictions against manually annotated reference labels, indicating whether a punchline was successfully retained in the summary. The outcome is visualized as a confusion matrix, offering an interpretable assessment of precision, recall, and overall fidelity in humor retention.

Figure 8 shows the confusion matrix for the best-performing configuration, namely the Attention-Augmented LSTM with Cohesion-Aware Decoding. The visualization reveals many True Positives (TP), indicating that the model successfully identified and retained valid punchlines in its summaries. Similarly, the True Negative (TN) quadrant confirms that non-humorous content was correctly excluded. Notably, the False Positive (FP) rate remains minimal, suggesting that the model seldom inserts sentences wrongly perceived as humorous. However, a moderate number of False Negatives (FN) is evident, highlighting the model's occasional failure to capture subtle or context-dependent humor, a persistent challenge in abstractive summarization tasks involving narrative texts.

Further interpretation of the matrix yields performance metrics that quantify punchline retention. The model achieved a precision of 0.87 and a recall of 0.79, resulting in an F1 score of 0.83. These values reflect a substantial improvement over baseline LSTM and transformer-only architectures, validating the effectiveness of integrating discourse-aware attention and cohesion-driven decoding. The results also substantiate the hypothesis that humor retention in summaries is not merely a matter of surface lexical features but a function of deeper narrative structures and contextual cue identification.

From a system design perspective, these findings highlight the utility of punchline-aware evaluation in the training and inference phases. Incorporating punchline prediction as an auxiliary task may enhance the model's semantic representations of humor. Furthermore, the diagnostic insights provided by the matrix, especially within the FN quadrant, offer guidance for future dataset enhancements and feature

engineering. Examples include augmenting the corpus with diverse humor styles or annotating rhetorical devices like irony, exaggeration, or incongruity, which often serve as punchline triggers.

In conclusion, this evaluation provides a nuanced dimension that complements lexical-based metrics by focusing on humor-specific retention. The confusion matrix framework empirically validates model performance and a strategic pathway for improving future abstractive summarization systems in humor-intensive domains. It reinforces the importance of evaluating what is summarized and how effectively critical rhetorical functions, such as humor, are preserved in the summarization process.

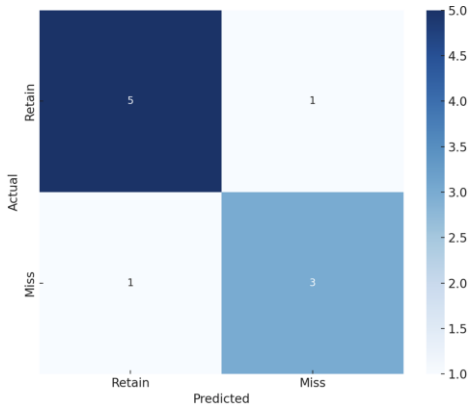


Fig 8. Confusion Matrix: Punchline Retention Evaluation

5. Conclusion

This study introduced a novel Attention-Augmented LSTM architecture with discourse-aware decoding to enhance the abstractive summarization of humorous narratives. By leveraging a unique dataset of over 10,000 Indonesian stand-up comedy transcripts enriched with behavioral annotations such as audience laughter, the model significantly improved the retention of rhetorical structures, particularly the interplay between setup and punchline. Empirical results confirmed that the integration of cohesion-first decoding and attention mechanisms contributed to higher ROUGE scores, better narrative coherence, and enhanced punchline retention—validated both quantitatively and through human evaluations.

Beyond lexical fidelity, the model exhibited strong alignment with audience responses, indicating its ability to learn what was said and how it was received. The findings also revealed that punchline density, narrative length, and performer style significantly influence summarization effectiveness, underscoring the importance of context-sensitive modeling. The cohesion-first approach effectively preserved thematic continuity and affective structure across varying discourse styles.

Future research should explore multilingual and cross-cultural extensions to assess model generalizability across different comedic traditions. Integrating multimodal inputs such as facial expressions, prosody, or visual stage cues could refine the model's sensitivity to humor delivery. Another promising direction lies in developing adaptive summarization strategies that dynamically adjust based on individual speaker profiles, audience demographics, or cultural context. These future studies will help build more inclusive, responsive, and rhetorically aware summarization systems tailored for complex, human-centered narratives like humor.

Acknowledgment

The authors would like to express their most profound appreciation to the research teams at the Faculty of Engineering and the Faculty of Letters, Universitas Negeri Malang, as well as the Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang, for their invaluable support and collaboration throughout this study. Special thanks are extended to the anonymous reviewers whose constructive feedback significantly improved the quality of this manuscript.

References

[1] A. D. Waluyaningtyas, “Utilization of Artificial Intelligence in Strengthening the Pancasila Student Profile Project Integrated with STEM,” vol. 5, no. 1, pp. 35–40, 2025.

[2] M. Kirmani, G. Kaur, and M. Mohd, “ShortMail: An email summarizer system,” *Softw. Impacts*, vol. 17, p. 100543, 2023, doi: https://doi.org/10.1016/j.simpa.2023.100543.

[3] Supriyono, A. P. Wibawa, Suyono, and F. Kurniawan, “A survey of text summarization: Techniques, evaluation and challenges,” *Nat. Lang. Process. J.*, vol. 7, no. March, p. 100070, 2024, doi: 10.1016/j.nlp.2024.100070.

[4] T. G. Altundogan and M. Karakose, “LSTM Encoder Decoder Based Text Highlight Abstraction Method Using Summaries Extracted by PageRank,” 2023, doi: 10.1109/IT57431.2023.10078652.

[5] M. Umer, I. Ashraf, A. Mehmood, S. Kumari, S. Ullah, and G. Sang Choi, “Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model,” *Comput. Intell.*, vol. 37, no. 1, pp. 409–434, 2021, doi: 10.1111/coin.12415.

[6] E. M. Saoudi, J. Jaafari, and S. J. Andaloussi, “Advancing human action recognition: A hybrid approach using attention-based LSTM and 3D CNN,” *Sci. African*, vol. 21, p. e01796, 2023, doi: https://doi.org/10.1016/j.sciaf.2023.e01796.

[7] R. Mirsa, M. Muhammad, E. Saputra, and I. Farhana, “Space Pattern of Samudera Pasai Sultanate,” *Int. J. Eng. Sci. Inf. Technol.*, vol. 1, no. 2, 2021, doi: 10.52088/ijesty.v1i2.120.

[8] A. Irfan Rifai, D. Fazadi Rafianda, M. Isradi, and A. Mufhidin, “Analysis Of Customer Satisfaction On The Application Of The Covid-19 Protocol At The Inter-City Bus Terminal,” *Int. J. Eng. Sci. Inf. Technol.*, vol. 1, no. 1, 2021, doi:

10.52088/ijesty.v1i1.107.
[9] R. Sepúlveda-Torres, M. Vicente, E. Saquete, E. Lloret, and M. Palomar, "Leveraging relevant summarized information and multi-layer classification to generalize the detection of misleading headlines," *Data Knowl. Eng.*, vol. 145, p. 102176, 2023, doi: https://doi.org/10.1016/j.datak.2023.102176.
[10] Supriyono, A. P. Wibawa, Suyono, and F. Kurniawan, "Advancements in natural language processing: Implications, challenges, and future directions," *Telemat. Informatics Reports*, vol. 16, p. 100173, Dec. 2024, doi: 10.1016/j.teler.2024.100173.
[11] A. F. U. R. Khilji *et al.*, "Multimodal text summarization with evaluation approaches," *Sadhana - Acad. Proc. Eng. Sci.*, vol. 48, no. 4, 2023, doi: 10.1007/s12046-023-02284-z.
[12] A. P. Wibawa, H. K. Fithri, I. A. E. Zaeni, and A. Nafalski, "Generating Javanese Stopwords List using K-means Clustering Algorithm," *Knowl. Eng. Data Sci.*, vol. 3, no. 2, p. 106, Dec. 2020, doi: 10.17977/um018v3i22020p106-111.
[13] T. Habib, A. Siregar, D. Abdullah, and L. Rosnita, "Plagiarism Detection Application for Computer Science Student Theses Using Cosine Similarity and Rabin-Karp," vol. 5, no. 1, pp. 185–194, 2025.
[14] Q. Zaman, "Supporting Application Fast Learning of Kitab Kuning for Santri ' Ula Using Natural Language Processing Methods," vol. 5, no. 1, pp. 278–289, 2025.
[15] M. Jiang, B. Lin, S. Wang, Y. Xu, W. Yu, and C. Zhu, "Summary and Future Directions," in *SpringerBriefs in computer science*, 2024, pp. 91–95.
[16] J. Kumar, R. Vashistha, R. Lal, and D. Somanir, "YouTube Transcript Summarizer," in *2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023*, Jul. 2023, pp. 1–4, doi: 10.1109/ICCCNT56998.2023.10308325.
[17] D. Pernes, "Towards End-to-end Speech-to-text Summarization," *arXiv.org*, vol. abs/2306.0, 2023, doi: 10.48550/arXiv.2306.05432.
[18] N. Vanetik, M. Litvak, and S. Krimberg, "Summarization of financial reports with TIBER," *Mach. Learn. with Appl.*, vol. 9, p. 100324, 2022, doi: https://doi.org/10.1016/j.mlwa.2022.100324.
[19] J. Guo, J. Liu, X. Liu, Y. Wan, and L. Li, "Summarizing source code with Heterogeneous Syntax Graph and dual position," *Inf. Process. Manag.*, vol. 60, no. 5, p. 103415, 2023, doi: https://doi.org/10.1016/j.ipm.2023.103415.
[20] F. A. Ahda, A. P. Wibawa, D. Dwi Prasetya, and D. Arbian Sulistyo, "Comparison of Adam Optimization and RMS prop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation," *JOIV Int. J. Informatics Vis.*, vol. 8, no. 1, p. 231, Mar. 2024, doi: 10.62527/joiv.8.1.1818.
[21] T. Widiyaningtyas, D. D. Prasetya, and H. W. Herwanto, "Time Loss Function-based Collaborative Filtering in Movie Recommender System," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 6, pp. 1021–1030, Dec. 2023, doi: 10.22266/ijies2023.1231.84.
[22] L. L. L. Zhang *et al.*, "Simulation of E-learning virtual interaction in Chinese language and literature multimedia teaching system based on video object tracking algorithm," *Expert Syst. Appl.*, vol. 10, no. 2, p. 102585, 2024, doi: https://doi.org/10.1016/j.jksuci.2020.08.007.
[23] D. Ramesh and S. K. Sanampudi, "Coherence-based automatic short answer scoring using sentence embedding," *Eur. J. Educ.*, vol. 59, no. 3, Sep. 2024, doi: 10.1111/ejed.12684.
[24] Z. Zhang, "Advancements and challenges in AI-driven language technologies: From natural language processing to language acquisition," *Applied and Computational Engineering*, vol. 57, no. 1. pp. 146–152, 2024, doi: 10.54254/2755-2721/57/20241325.
[25] Rishu and V. Kukreja, "Comic exploration and Insights: Recent trends in LDA-Based recognition studies," *Expert Syst. Appl.*, vol. 255, p. 124732, 2024, doi: https://doi.org/10.1016/j.eswa.2024.124732.
[26] D. Alahmadi, A. Wali, and S. Alzahrani, "TAAM: Topic-aware abstractive arabic text summarisation using deep recurrent neural networks," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, Part A, pp. 2651–2665, 2022, doi: https://doi.org/10.1016/j.jksuci.2022.03.026.