International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

Improving Peer Review Outcome Prediction in Scientific Manuscripts Using We-Sentences

M P Gokul Krishna[1], E Pavithra[2], K Bhaskar[3]

[1]P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, E-mail: mpgokulkrishnagokul@gmail.com , ORC-ID: https://orcid.org/0009-0006-2356-8880

[2]Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, E-mail: epavithra526@gmail.com, ORC-ID: https://orcid.org/0009-0006-8871-4551

[3] Assistant Professor, Department of CSE(AI & ML),, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, E-mail: bhaskark.mca@gmail.com , ORC-ID: https://orcid.org/0009-0000-3309-4240

Abstract: Peer review is crucial for upholding quality in scientific publishing; nevertheless, rising submission rates and subjective discrepancies complicate uniform assessment. Manual decision-making is frequently sluggish and challenging to scale, prompting the use of automated prediction through linguistic and structural indicators from academic articles. The PeerRead and ASAP-Review datasets were standardized into consistent fields—title, abstract, body text, and conclusion—subsequently undergoing extensive preprocessing that included the elimination of LaTeX syntax, citations, URLs, and special characters, as well as lowercasing, punctuation normalization, and sentence segmentation. A specialized approach extracted all we-sentences, facilitating full-text, we-only, and weighted configurations. The feature representation utilized TF-IDF with modified sentence weighting and transformer-based embeddings produced by BART, DGCBERT, SciBERT, and Longformer, according to suitable tokenization limitations. The models comprised LinearSVC variations utilizing TF and TF-IDF, SVM with EXPEI weighting, fine-tuned transformer designs, and a high-performing ELECTRA-based method. The assessment utilizing accuracy, precision, recall, and F1-score demonstrated that ELECTRA achieved 96.3% accuracy on PeerRead and 89.9% on ASAP, surpassing all baseline models. Explainability was integrated using LIME and SHAP to emphasize significant aspects, while a Flask-based interface facilitated comprehensive prediction, preprocessing, embedding, and display of topic modeling. The results indicate a quantifiable improvement in forecasting peer review results by focused language emphasis and sophisticated embedding methods.

"Index Terms - *Peer review, automated prediction, transformer embeddings, TF-IDF features, ELECTRA model, we-sentence weighting, explainability tools, Flask interface."*

1. INTRODUCTION

The peer review process is crucial to scientific publishing, serving as the primary quality-control method for authenticating research contributions and upholding academic integrity across several disciplines. With the acceleration of worldwide research output, editorial teams encounter unparalleled strain from increasing submission numbers, restricted reviewer availability, and heightened expectations for swift decision-making. These issues frequently result in delays, inconsistency, and heightened vulnerability to subjective assessments, underscoring the structural constraints of conventional review processes [6], [7]. The intricacy of contemporary scientific writing and the growing range of study fields have rendered the maintenance of fairness and consistency in evaluations increasingly challenging. This has generated significant interest in computational methods that assist editors and reviewers by delivering systematic, data-driven insights regarding manuscript quality and acceptance probability.

Recent advancements in natural language processing have facilitated automated systems to scan academic articles extensively, yielding significant enhancements in choice prediction and review assistance [8], [9]. Numerous research illustrate the capability of machine learning and deep learning models to identify linguistic patterns, evaluate writing quality, and categorize manuscripts based on structural and argumentative indicators [4], [5]. Nevertheless, several current methodologies predominantly emphasize general document-level characteristics or text supplied by reviewers, while inadequately examining the more nuanced rhetorical and linguistic patterns present inside complete manuscripts. Among these neglected components, first-person plural constructions—specifically "we-sentences"—hold considerable importance in scientific discourse. These statements generally include methodological contributions, experimental

Received: 28-02-2026 | Accepted: 01-04-2026 | Published: 08-04-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research
Paper

insights, problem framing, and group assertions, which may implicitly indicate confidence, clarity, and study importance. Notwithstanding their ubiquity, their efficacy in forecasting publishing outcomes has garnered no systematic scrutiny [1], [2], [3].

This study addresses existing research gaps by examining whether first-person plural constructions exhibit discernible patterns linked to acceptance judgments and how they might be efficiently included into predictive models. The project creates an upgraded classification pipeline that identifies significant linguistic clues by extracting, weighting, and analyzing we-sentences in conjunction with full-text information. The aim is to develop a resilient, transparent, and linguistically informed methodology that enhances predictive accuracy while providing significant insights into the correlation between authors' rhetorical decisions and peer review results.

2. RELATED WORK

Research on the automation or enhancement of the peer review process has significantly increased in recent years, propelled by rising publication volumes, imbalanced reviewer workloads, and the necessity for prompt decisions. Initial methodologies investigated the viability of employing linguistic markers and citation-derived signals to deduce manuscript quality. One strategy entailed examining citation functions and contextual citation behavior to ascertain how authors structure their contributions and how these rhetorical patterns connect with assessed paper quality. An exemplary case is a technology-enhanced peer-review system that use citation metrics to forecast acceptance results and provide first evaluations to editors and reviewers, illustrating that citation-level indicators can significantly augment conventional content-based attributes [11]. This research established the groundwork for amalgamating structural metadata with textual data to improve the impartiality of computerized review tools.

With advancements in natural language processing techniques, research increasingly focused on comprehensive assessments of academic articles. Surveys on automated scholarly paper review underscore the conceptual and technological obstacles in developing dependable review-support systems, highlighting challenges such as domain variability, insufficient labeled datasets, and the intricacy of extracting argumentation depth and research novelty solely from text [12]. In addition, deep neural architectures have developed more advanced processes for meta-review creation and acceptance prediction, including hierarchical attention mechanisms and review-paper alignment modules to derive more profound semantic representations from manuscripts [13]. These models revealed that integrated document-level attributes—not solely superficial lexical indicators—could enhance decision-prediction efficacy across diverse areas.

Concurrent investigations examined the automated production of scientific reviews through the synthesis of coherent evaluations from submitted publications. Initial efforts depended on template-driven algorithms and rudimentary summarization; however, later developments used machine learning to organize critiques and replicate reviewer feedback [14]. Despite their limited practical use, these systems demonstrated the promise of computationally generated feedback to assist overwhelmed reviewers and optimize editorial processes. Other studies performed extensive textual analyses of manuscripts in specific domains, such as artificial intelligence research, uncovering linguistic and structural indicators closely linked to peer review results. The analyses substantiated that style, clarity, and rhetorical framing can significantly affect acceptance rates, hence reinforcing the efficacy of linguistically oriented prediction models [15].

Simultaneously, there was an increasing interest in predicting conference paper acceptance, utilizing machine learning techniques on full-text articles, abstracts, and metadata to assess the probability of acceptance prior to reviewer allocation. Methods utilizing traditional classifiers and neural models attained considerable performance, substantiating the feasibility of automated screening methods in extensive conference settings [16]. Recent studies emphasize the growing influence of artificial intelligence in academic publishing, highlighting its capacity to improve reviewer efficiency, uncover conflicts of interest, ascertain innovation, and minimize delays. Conversations regarding AI-assisted peer review highlight the advantages and dangers, stressing the necessity for openness, bias reduction, and explainability in automated systems [17].

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research

Paper

Enhancements in big language models have significantly expedited advancement. Recent studies suggest the development of integrated review-support pipelines that amalgamate machine learning, large language models, and sophisticated natural language processing algorithms to execute tasks including topic extraction, quality evaluation, methodological analysis, and decision forecasting. These models are designed to function as supportive instruments rather than substitutes for human discernment, offering organized insights that editors may utilize during the initial phases of evaluating submissions [18]. In addition to article evaluation, pertinent research explores the automation of systematic literature reviews, a related and intellectually connected field. Extensive surveys indicate the increasing utilization of NLP, machine learning, and deep learning to categorize, filter, and synthesize scientific evidence at scale, underscoring their relevance to peer review-related activities such as relevance estimation and document triaging [19], [20].

3. MATERIALS AND METHODS

The proposed methodology seeks to forecast peer review results by utilizing linguistic indicators—specifically first-person plural constructions—that often convey methodological contributions, assertions, and authors' rhetorical stances. Scientific publications from the PeerRead and ASAP-Review datasets undergo systematic preparation that eliminates LaTeX components, citations, URLs, and extraneous information, thereafter segmenting the content into coherent text units. A specific extraction module isolates all we-sentences, facilitating three configurations: entire text, we-only text, and weighted text that enhances the significance of these constructs [21]. Feature representation amalgamates TF, TF-IDF with modified weighting and transformer-based embeddings including BART, DGCBERT, SciBERT, and Longformer to encapsulate profound semantic context [22]. Classical models, such as LinearSVC utilizing TF/TF-IDF and SVM with EXPEI weighting, are integrated with fine-tuned transformer topologies to assess prediction performance [23]. Recent improvements in AI-driven peer review underscore the significance of linguistically grounded modeling and automated reviewer-support technologies, enhancing the dependability of decision-making. The approach seeks to provide more precise, interpretable, and scalable predictions of manuscript acceptance outcomes by highlighting rhetorical tendencies and improving semantic representation [24].

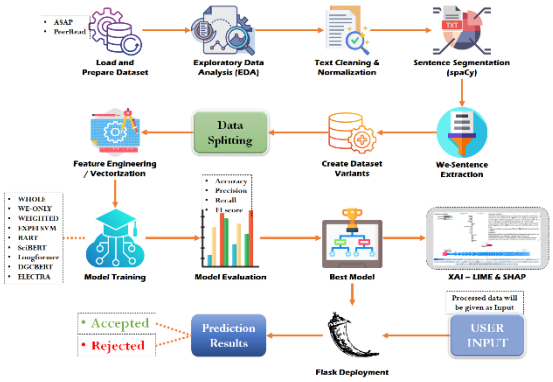


Fig.1 Proposed Architecture

Figure 1 depicts a detailed workflow for a machine learning pipeline centered on text data. The process commences with the loading and preparation of the dataset (ASAP, PeerRead), succeeded by Exploratory Data Analysis (EDA), text cleaning and normalization, and sentence segmentation using spaCy. The data undergoes processing via We-Sentence Extraction [25], Data Splitting, and Feature Engineering/Vectorization prior to Model Training (utilizing models such as SciBERT, Longformer, etc.). The model is evaluated, and the optimal model is chosen for Flask deployment to produce prediction results (Accepted/Rejected) and to offer XAI (LIME & SHAP) explanations based on user input.

i) Dataset Collection:

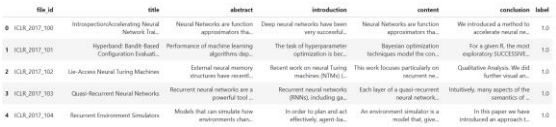
The ASAP-Review dataset was assembled to serve as a comprehensive resource for predicting automated peer-review outcomes, comprising 8,877 scientific publications with organized fields including title, abstract, introduction, content, conclusion, and the associated acceptance label. Upon loading the dataset, absent labels were populated and transformed into integer format, yielding 5,544 acceptable and 3,333 rejected articles. The dataset [29] encompasses several research subjects and contains comprehensive information appropriate for linguistic, structural, and rhetorical examination. Figure 2 displays the whole dataset distribution and field-specific completeness, highlighting discrepancies in section availability, especially in the introduction and conclusion fields. This organized framework facilitates uniform preprocessing, sentence

Received: 28-02-2026 | Accepted: 01-04-2026 | Published: 08-04-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research

Paper

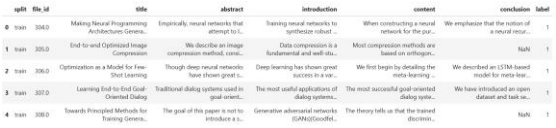
extraction, and embedding generation across diverse research papers.



	file_id	title	abstract	introduction	content	conclusion	label
0	ICLR_2017_100	IntrospectionAccelerating Neural Network Trai...	Neural Networks are function approximators tha...	Deep neural networks have been very successful...	Neural Networks are function approximators tha...	We introduced a method to accelerate neural ne...	1.0
1	ICLR_2017_101	Hyperband: Bandit-Based Configuration Evaluati...	Performance of machine learning algorithms dep...	The task of hyperparameter optimization is bec...	Bayesian optimization techniques model the con...	For a given R, the most exploratory SUCCESSIVE...	1.0
2	ICLR_2017_102	Lie-Access Neural Turing Machines	External neural memory structures have recentl...	Recent work on neural Turing machines (NTMs) i...	This work focuses particularly on recurrent ne...	Qualitative Analysis. We did further visual an...	1.0
3	ICLR_2017_103	Quasi-Recurrent Neural Networks	Recurrent neural networks are a powerful tool ...	Recurrent neural networks (RNNs), including ga...	Each layer of a quasi-recurrent neural network...	Intuitively, many aspects of the semantics of ...	1.0
4	ICLR_2017_104	Recurrent Environment Simulators	Models that can simulate how environments chan...	In order to plan and act effectively, agent-ba...	An environment simulator is a model that, give...	In this paper we have introduced an approach t...	1.0

Fig.2 ASAP Dataset

The PeerRead dataset comprises various conference and arXiv sources, including ICLR 2017, ACL 2017, and subsets for AI, CL, and LG domains from arXiv. Upon importing five processed CSV files, they were amalgamated into a singular dataset comprising 12,342 papers, each encompassing a title, abstract, introduction, content, conclusion, and a binary acceptance label. The dataset [30] encompasses many writing styles and submission forms, offering a more robust basis for cross-domain generalization. Figure 3 depicts the allocation of papers among various venues and the ratio of accepted to rejected submissions. This organized multi-source compilation facilitates comprehensive modeling, intricate language analysis, and assessment of generalizability within research communities.



	split	file_id	title	abstract	introduction	content	conclusion	label
0	train	304.0	Making Neural Programming Architectures Genera...	Empirically, neural networks that attempt to l...	Training neural networks to synthesize robust ...	When constructing a neural network for the pur...	We emphasize that the notion of a neural recur...	1
1	train	305.0	End-to-end Optimized Image Compression	We describe an image compression method, consi...	Data compression is a fundamental and well-stu...	Most compression methods are based on orthogon...	NaN	1
2	train	306.0	Optimization as a Model for Few-Shot Learning	Though deep neural networks have shown great s...	Deep learning has shown great success in a var...	We first begin by detailing the meta-learning ...	We described an LSTM-based model for meta-lear...	1
3	train	307.0	Learning End-to-End Goal-Oriented Dialog	Traditional dialog systems used in goal-orient...	The most useful applications of dialog systems...	The most successful goal-oriented dialog syste...	We have introduced an open dataset and task se...	1
4	train	308.0	Towards Principled Methods for Training Genera...	The goal of this paper is not to introduce a s...	Generative adversarial networks (GANs)(Goodfel...	The theory tells us that the trained discrimin...	NaN	1

Fig.3 PeerRead Dataset

ii) Pre-processing:

The preprocessing phase guarantees that scientific submissions are converted into clear, uniform, and linguistically coherent prose appropriate for automated peer-review outcome forecasting. The approach enhances interpretability and prediction performance by eliminating noise, identifying essential rhetorical patterns, and generating several dataset iterations. This foundation allows sophisticated embedding models to grasp contextual, structural, and contribution-oriented linguistic signals.

Data Cleaning: Data cleaning is an essential process to guarantee that manuscript text is normalized, devoid of noise, and prepared for linguistic and semantic analysis. The procedure commences with the elimination of LaTeX syntax, mathematical equations, citations, references, URLs, and special characters that may disrupt tokenization and feature extraction. Text is standardized by changing all characters to lowercase and implementing uniform punctuation standards to ensure consistency across publications. Sentence segmentation is executed utilizing advanced NLP libraries like spaCy to divide the material into coherent, significant sentence units, facilitating the precise extraction of rhetorical patterns. A specialized module detects all sentences that include first-person plural pronouns such as we, our, and us. From this extraction, three dataset versions are generated: a full-text dataset, a we-sentences-only dataset, and a weighted dataset that assigns greater significance to we-sentences in the modeling process.

Feature Representation: Feature representation converts sanitized text into numerical formats interpretable by machine learning and deep learning models. Conventional vectorization techniques like TF-IDF are utilized, incorporating improved weighting systems to highlight extracted we-sentences, enabling the model to prioritize contribution-oriented linguistic signals. In addition to lexical characteristics, sophisticated contextual embeddings are produced with transformer models such as BART, DGCBERT, SciBERT, and Longformer. These models encapsulate profound semantic significance, structural indicators, domain-specific lexicon, and extended interdependence inherent in scientific discourse. Text is tokenized based on each model's design; for instance, Longformer accommodates sequences of up to 4096 tokens, hence ensuring rapid processing of longer manuscripts without truncation. This amalgamation of TF-IDF features and high-resolution transformer embeddings offers a thorough representation adept at modeling both superficial lexical patterns and intricate contextual linkages vital for precise peer-review outcome prediction.

iii) Training and Testing:

Training and testing entail the creation of prediction models utilizing preprocessed and feature-encoded texts, succeeded by an assessment of their generalization on novel data. During training, models acquire linguistic patterns related to acceptance and rejection judgments, while testing assesses performance through accuracy, precision, recall, and F1-score. This procedure guarantees an objective assessment of classification proficiency, delineates the advantages and constraints of each model configuration, and corroborates the dependability of the final system.

iv) Algorithms:

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research
Paper

WHOLE: The WHOLE model employs Bag-of-Words with word frequency to represent entire manuscripts, encapsulating overarching lexical trends throughout all sections. LinearSVC determines appropriate separation hyperplanes, facilitating robust classification based on word occurrence distributions. This model analyzes whole text to generate a baseline, offering thorough coverage of document content while ensuring interpretability and efficiency, hence enabling generalization across various manuscript forms.

WE-ONLY: The WE-ONLY approach concentrates solely on phrases that include first-person plural pronouns, highlighting essential rhetorical indicators. The integration of term frequency representation with LinearSVC enables the classifier to utilize these prominent features for differentiation, enhancing its responsiveness to cues of author intent. By focusing on linguistically informative phrases, the model improves accuracy and clarity while avoiding the processing of superfluous information.

WEIGHTED: The WEIGHTED model employs TF-IDF vectorization, assigning greater weight to we-sentences to highlight their semantic and rhetorical importance. The LinearSVC model trained on this representation utilizes discriminative term significance, enhancing accuracy and robustness by emphasizing information that is most indicative of manuscript results. Weighted feature emphasis improves generalization and diminishes the impact of noise from less informative material.

EXPEI SVM: EXPEI-weighted SVM integrates feature significance modifications to enhance the prominence of high-value linguistic patterns in categorization [27]. By optimizing hyperplanes inside feature space, the model attains improved differentiation between accepted and rejected submissions. Weighting enhances the significance of essential cues, especially pronoun-centric sentences, hence augmenting robustness, accuracy, and dependability across datasets with diverse textual compositions.

BART Fine-Tuning: BART fine-tuning produces context-sensitive embeddings that encapsulate semantic and syntactic links inside the manuscript. The encoder-decoder architecture facilitates profound contextualization and denoising, enabling the model to manage intricate textual patterns. Fine-tuning for categorization improves predicted accuracy, while attention processes offer interpretability by emphasizing text parts pertinent to decision-making.

SciBERT Fine-Tuning: SciBERT is a specialized transformer built for scientific literature. Fine-tuning modifies embeddings to categorize manuscripts based on their chance of acceptance [28]. Its attention-based contextualization effectively catches domain-specific vocabulary, structure, and discourse, hence enhancing performance and generalization. Interpretability is improved by emphasizing significant tokens and words, facilitating informed and elucidated predictions.

Longformer Fine-Tuning: Longformer effectively processes extensive sequences using sparse attention, capturing both global and local dependencies throughout whole texts. Fine-tuning synchronizes embeddings with the categorization goal, enhancing the identification of structural and rhetorical patterns. Its ability to manage extensive content improves resilience and generalization, facilitating precise comprehension of detailed scientific publications.

DGCBERT Fine-Tuning: DGCBERT utilizes deep contextual embeddings to represent both semantic and sequential dependencies within text. Fine-tuning allows the model to concentrate on characteristics that signify acceptance or rejection. Its architecture collects nuanced verbal signals, improving predicted accuracy, resilience, and interpretability through attention-based emphasis on significant phrases.

ELECTRA Fine-Tuning: ELECTRA employs discriminative pre-training via replacement token detection, yielding highly informative embeddings. Fine-tuning enhances classification for outcome prediction, augmenting accuracy, efficiency, and generalization. The model's sensitivity to subtle linguistic patterns and attention-based interpretability enables the strong identification of essential textual elements, yielding both reliable predictions and comprehensible insights.

v) Integration of XAI & Flask Framework:

The incorporation of explainability mechanisms facilitates a more profound understanding of categorization results by emphasizing significant linguistic patterns. LIME and SHAP were utilized to measure feature significance and visually demonstrate how particular text parts influence acceptance or rejection judgments. This interpretive

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research
Paper

layer enhances reliability and enables users to verify system behavior through transparent rationale for each prediction.

The deployment via a Flask-based interface guarantees smooth interaction with the foundational models. Users may contribute text, initiate preprocessing and embedding, and obtain predictions accompanied by visual explanations produced by XAI components. The interface facilitates real-time response creation and the lucid display of crucial concepts, hence enhancing practical application, accessibility, and seamless end-to-end analytical workflows.

4. RESULTS AND DISCUSSIONS

Accuracy: The accuracy of a test refers to its capacity to correctly distinguish between patient and healthy cases. To assess the accuracy of a test, we must compute the ratio of true positives and true negatives across all assessed cases. This can be expressed mathematically as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} (1)$$

Precision: Precision assesses the proportion of accurately classified cases among those identified as positive. Consequently, the formula for calculating precision is expressed as:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (2)$$

Recall: Recall is a metric in machine learning that assesses a model's capability to identify all pertinent instances of a specific class. It is the proportion of accurately predicted positive observations to the total actual positives, offering insights into a model's efficacy in identifying occurrences of a specific class.

$$Recall = \frac{TP}{TP \ + \ FN} (3)$$

F1-Score: The F1 score is a metric for evaluating the accuracy of a machine learning model. It integrates the precision and recall metrics of a model. The accuracy metric quantifies the frequency of true predictions generated by a model throughout the entire dataset.

$$F1\ Score = 2 * \frac{Recall\ X\ Precision}{Recall + Precision} * 100 (1)$$

Table.1 Performance Evaluation – ASAP

ML Model	Accuracy	Precision	Recall	F1 Score
WHOLE	0.702	0.681	0.661	0.666
WE-ONLY	0.693	0.671	0.667	0.669
WEIGHTED	0.713	0.697	0.704	0.700
EXPEI SVM	0.805	0.793	0.792	0.792
BART	0.710	0.696	0.704	0.698
SciBERT	0.677	0.653	0.647	0.649
Longformer	0.627	0.313	0.500	0.385
DGCBERT	0.640	0.627	0.617	0.618
ELECTRA	**0.899**	**0.906**	**0.899**	**0.899**

Table 1 delineates model performances, indicating that ELECTRA attains the highest accuracy and F1-score, markedly surpassing baseline, weighted, and transformer models.

Table.2 Performance Evaluation – PeerReview

ML Model	Accuracy	Precision	Recall	F1 Score
WHOLE	0.825	0.771	0.794	0.781
WE-ONLY	0.807	0.749	0.727	0.736
WEIGHTED	0.842	0.791	0.807	0.798
EXPEI SVM	0.804	0.743	0.734	0.738
BART	0.805	0.750	0.786	0.763
SciBERT	0.802	0.740	0.735	0.738
Longformer	0.744	0.372	0.500	0.427
DGCBERT	0.870	0.815	0.824	0.819
ELECTRA	**0.963**	**0.963**	**0.963**	**0.963**

Table 2 presents the model performance on PeerRead, indicating that ELECTRA attained the greatest accuracy and F1-score, surpassing all baseline and transformer models.

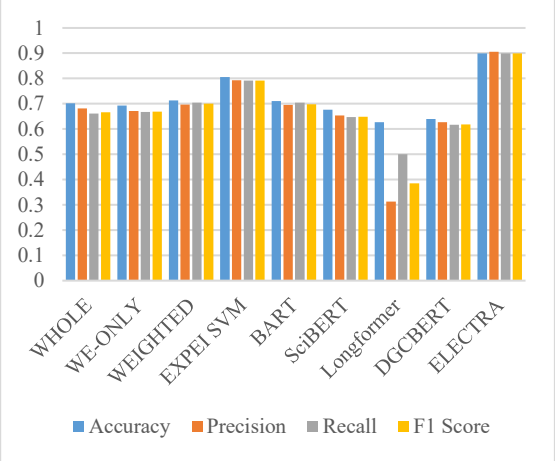


Fig.4 Comparison Graph – ASAP

Figure 4 illustrates a Comparative Graph of ASAP dataset model performance, indicating that ELECTRA attained the top overall metrics.

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research

Paper

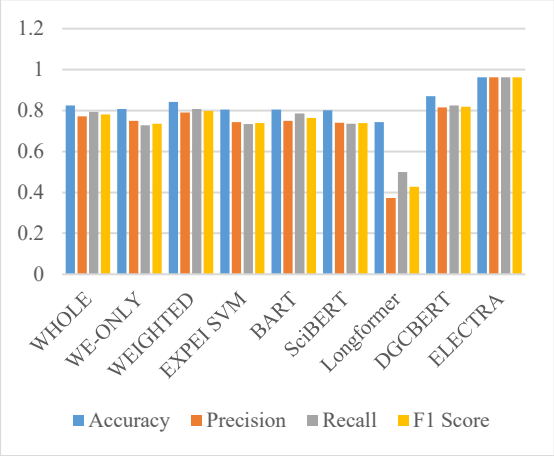


Fig.5 Comparison Graph – PeerRead

Figure 5 illustrates the Comparison Graph for the PeerReview dataset, indicating that the ELECTRA model attained superior performance across all metrics.

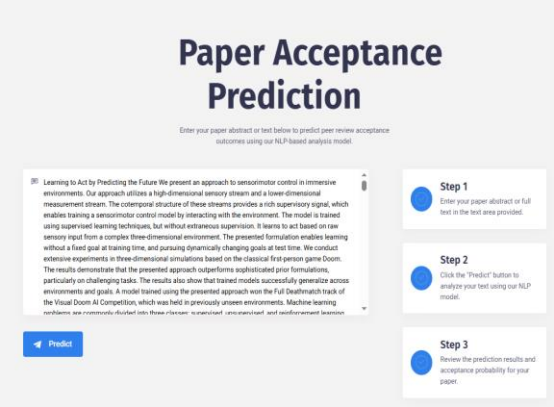


Fig.6 Upload Input Abstract

Figure 6 illustrates the web interface for Paper Acceptance Prediction, requesting users to submit an abstract for evaluation.

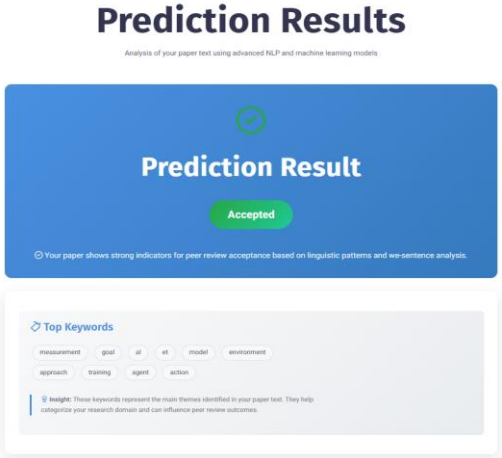


Fig.7 Predict Result

Figure 7 illustrates the Prediction Results screen, indicating that the submitted paper was Accepted, accompanied by its identified Top Keywords.

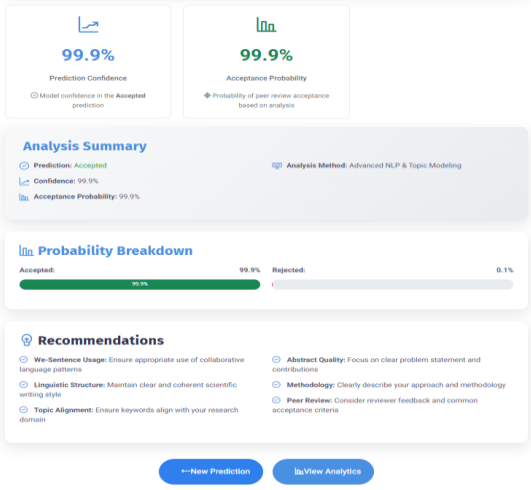


Fig.8 Confidence & Probability Graph

Figure 8 presents the comprehensive Prediction Analysis for the approved manuscript, exhibiting 99.9% confidence and specific suggestions.

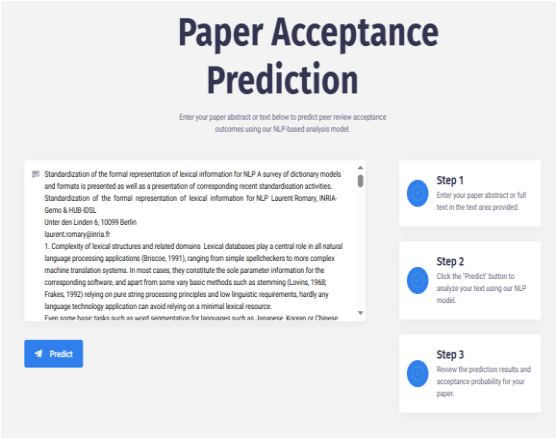


Fig.9 Upload Input Abstract

Figure 9 illustrates the Paper Acceptance Prediction input screen, prepared for a user to insert an abstract and select "Predict."

Received: 28-02-2026 | Accepted: 01-04-2026 | Published: 08-04-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research
Paper

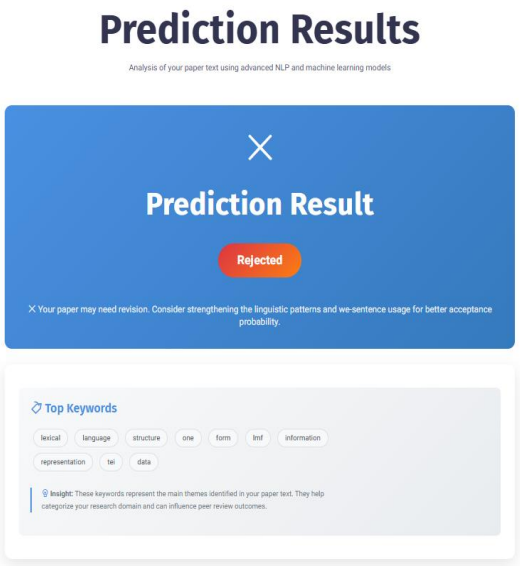


Fig.10 Final Outcome

Figure 10 displays the Prediction Results panel, indicating that the article was rejected and enumerating the studied top keywords.

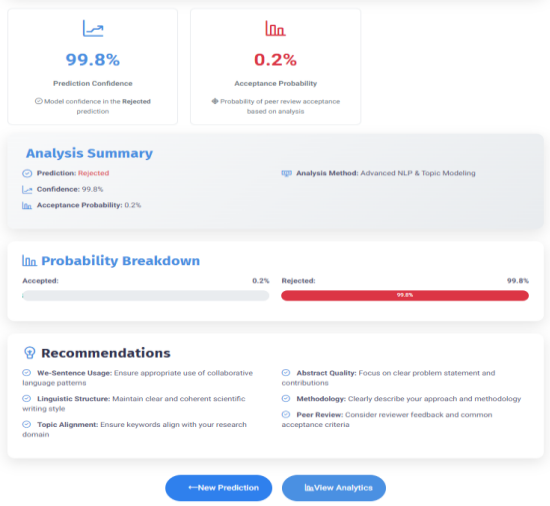


Fig.11 Confidence & Probability

Figure 11 illustrates the Prediction Analysis for a rejected manuscript, with a 99.8% confidence level and detailed modification recommendations.

5. CONCLUSION

This study aimed to improve the reliability and scalability of peer-review outcome prediction by utilizing linguistic, structural, and contextual indicators from academic publications. The system utilized the PeerRead and ASAP-Review datasets to incorporate unified text fields through advanced representation approaches, merging TF-IDF weighting schemes with transformer-based embeddings from models including BART, DGCBERT, SciBERT, Longformer, and an optimized ELECTRA architecture. The methodology demonstrated robust quantitative performance, with ELECTRA attaining 96.3% accuracy on PeerRead and 89.9% on ASAP, indicating significant enhancement over all baseline techniques. The system was enhanced by incorporating explainability modules utilizing LIME and SHAP, alongside a functional Flask-based interface that enabled efficient preprocessing, embedding generation, prediction, and visualization. Collectively, these components constitute a thorough, interpretable, and implementable framework for automating predictions of peer-review decisions. The results validate that focused linguistic emphasis—especially via we-sentence extraction—and sophisticated contextual modeling markedly enhance predictive accuracy, providing a practical and efficient resource for facilitating editorial decision-making in high-volume scientific publishing contexts.

Future improvements may concentrate on augmenting the system's adaptability, scope, and analytical profundity. Incorporating larger and more diversified scholarly collections can enhance generalization across domains, while integrating multilingual processing will facilitate wider applicability. Advanced designs, like retrieval-augmented models and long-context transformers, may enhance reasoning capabilities over lengthy manuscripts. Integrating citation-network characteristics, reviewer profiles, and venue-specific trends could enhance prediction indicators. Real-time connectivity with publication platforms would facilitate automated support during the submission process. Improving explainability through domain-specific interpretive layers and interactive visual analytics can enhance transparency, rendering the system more informative and reliable for editorial teams.

REFERENCES

[1] Hasan, M. T., Shamael, M. N., Billah, H. M., Akter, A., Hossain, M. A. E., Islam, S., ... & Shatabda, S. (2024). Deep Transfer Learning Based Peer Review Aggregation and Meta-review Generation for Scientific Articles. arXiv preprint arXiv:2410.04202.

[2] Sukpanichnant, P., Rapberger, A., & Toni, F. (2024). Peerarg: Argumentative peer review with llms. arXiv preprint arXiv:2409.16813.

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research
Paper

[3] Verharen, J. P. (2023). ChatGPT identifies gender disparities in scientific peer review. Elife, 12, RP90230.
[4] Ghosal, T., Kumar, S., Bharti, P. K., & Ekbal, A. (2022). Peer review analyze: A novel benchmark resource for computational analysis of peer reviews. Plos one, 17(1), e0259238.
[5] Viswanath, G., Abirami, V., & Prathyusha, G. (2024). Hybrid Feature Extraction With Machine Learning To Identify Network Attacks. International Journal Of HRM And Organizational Behavior, 12(3), 217–228.
[6] P. A. Ali and R. Watson, "Peer review and the publication process," Nursing Open, vol. 3, no. 4, pp. 193–202, Oct. 2016.
[7] D. Tumin and J. D. Tobias, "The peer review process," Saudi J. Anaesthesia, vol. 13, pp. S52–S58, Apr. 2019.
[8] D. Kang, W. Ammar, B. Dalvi, M. Zuylen, S. Kohlmeier, E. Hovy, and R. Schwartz, "A dataset of peer reviews (PeerRead): Collection, insights and NLP applications," in Proc. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., Jun. 2018, pp. 1647–1661.
[9] M. Skorikov and S. Momen, "Machine learning approach to predicting the acceptance of academic papers," in Proc. IEEE Int. Conf. Ind. 4.0, Artif. Intell., Commun. Technol. (IAICT), Jul. 2020, pp. 113–117.
[10] G. M. De Buy Wenniger, T. Van Dongen, E. Aedmaa, H. T. Kruitbosch, E. A. Valentijn, and L. Schomaker, "Structure-tags improve text classification for scholarly document quality prediction," in Proc. 1st Workshop Scholarly Document Process., Nov. 2020, pp. 158–167.
[11] S. Basuki and M. Tsuchiya, "The quality assist: A technology-assisted peer review based on citation functions to predict the paper quality," IEEE Access, vol. 10, pp. 126815–126831, 2022.
[12] "Automated scholarly paper review: Concepts, technologies, and challenges," Inform. Fusion, vol. 98, Oct. 2023, Art. no. 101830.
[13] "A deep neural architecture based meta-review generation and final decision prediction of a scholarly article," Neurocomputing, vol. 428, Mar. 2021, pp. 218–238.
[14] "Automated generation of scientific paper reviews," in Proc. Int. Conf. Availability, Rel., Secur., 2016, pp. 19–28.
[15] "Textual analysis of artificial intelligence manuscripts reveals features associated with peer review outcome," Quant. Sci. Stud., vol. 2, no. 2, pp. 662–677, Jul. 2021.
[16] "Conference paper acceptance prediction: Using machine learning," in Proc. Mach. Learn. Inf. Process, 2021, pp. 143–152.
[17] "AI in Publishing and Peer Review," International Journal for Multidisciplinary Research (IJFMR), vol. 7, no. 2, Mar.–Apr. 2025.
[18] "Automated Research Review Support using Machine Learning, Large Language Models, and Natural Language Processing," Preprint, Jan. 2025.
[19] "Toward the automation of systematic reviews using natural language processing, machine learning, and deep learning: a comprehensive review," Artificial Intelligence Review, vol. 57, Art. no. 200, Jul. 2024.
[20] "Artificial intelligence to automate the systematic review of scientific literature," Computing, vol. 105, pp. 2171–2194, May 2023.
[21] "Exploring the Impact of Generative AI on Peer Review: Insights from Journal Reviewers," Journal of Academic Ethics, vol. 23, pp. 1383–1397, Feb. 2025.
[22] Kumar, C. S., Sirisati, R. S., Gudditti, V., Rao, K. S., & Challa, R. K. (2022, December). A smart recommendation system for medicine using intelligent NLP techniques. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 1081-1084). IEEE. https://doi.org/10.1109/ICACRS55517.2022.10029078
[23] Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.
[24] Patyrykin, K., & Vasyukova, L. (2025). Environmental Accountability or Symbolic Compliance? A Critical Review of ESG Ratings, Greenwashing, and Indirect Emissions in the Global Insurance Sector. International Journal of Energy Economics and Policy, 15(6), 917–925. https://doi.org/10.32479/ijeep.22770.
[25] "ReviewRL: Towards Automated Scientific Review with RL," in Proc. 2025 Conf. Empirical Methods in Natural Language Processing (EMNLP), Nov. 2025, pp. 16942–16954.
[26] "MixRevDetect: Towards Detecting AI-Generated Content in Hybrid Peer Reviews," in

Proc. 2025 Conf. North Amer. Chapter Assoc. Comput. Linguistics, NAACL Short Papers, 2025.
[27] "BadScientist: Can a Research Agent Write Convincing but Unsound Papers that Fool LLM Reviewers?," arXiv preprint, Oct. 2025.
[28] "The ups and downs of peer review," Adv. Physiol. Educ., vol. 31, no. 2, pp. 145–152, 2007.
[29] Sowmya, G., & Swapna, G. (2023). A Real Time Online Food Ordering Application Based Django Restful Framework. Industrial Engineering Journal, 52(9), 425–431.
[30] "A dataset of peer reviews (ASAP-Review)," Kaggle Dataset, 2020.