



JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY

<https://e-journal.uum.edu.my/index.php/jict>

How to cite this article:

Wahyono, A., & Ku-Mahamud, K. R. (2026). Computational clickbait detection: A PSALSAR-based systematic literature review of methodological trajectories and benchmarks (2016–2025). *Journal of Information and Communication Technology*, 25(4), 21-52. <https://doi.org/10.32890/jict2026.25.4.2>

Computational Clickbait Detection: A PSALSAR-based Systematic Literature Review of Methodological Trajectories and Benchmarks (2016–2025)

^{*1}Ari Wahyono & ²Ku Ruhana Ku-Mahamud

^{1&2}Faculty of Business and Information Technology,
Universiti Muhammadiyah Malaysia, Malaysia

¹Faculty of Engineering, Universitas Tiga Serangkai, Indonesia

²School of Computing, Universiti Utara Malaysia, Malaysia

^{*1}ariwahyono@tsu.ac.id

²ruhana@uum.edu.my

^{*}Corresponding author

Received: 8/5/2026

Revised: 6/8/2026

Accepted: 7/9/2026

Published: 31/10/2026

ABSTRACT

Clickbait headlines have become a persistent issue in online media, where attention often takes priority over informational accuracy. While numerous computational approaches have been proposed to address this issue, the field remains fragmented across datasets, modelling strategies, and linguistic representations. This study presents a systematic literature review of clickbait detection research published between 2016 and 2025, following the Protocol, Search, Appraisal, Synthesis, Analysis, and Report framework to ensure a structured and reproducible review process. A total of 646 records were initially identified from Scopus, IEEE Xplore, Web of Science, and SpringerLink, from which 43 peer-reviewed journal articles were selected for in-depth analysis. Across these studies, a clear shift is evident from traditional feature-based machine learning toward deep learning and transformer-based models. Despite this progression, linguistic cues remain central, either through explicit feature engineering or implicitly learned representations. However, the review reveals a persistent gap: linguistic feature modelling and embedding-based approaches are often developed in isolation. To address this limitation, this study proposes a conceptual framework that integrates language-aware preprocessing, explicit linguistic features, and hybrid embedding strategies within a unified framework. By synthesising these components, the framework offers a more comprehensive perspective on clickbait detection, positioning it as an integrated linguistic and contextual modelling task rather than a purely classification problem. The findings and proposed framework provide a foundation for developing more robust, interpretable, and adaptable clickbait detection systems, particularly in multilingual and low-resource settings.

Keywords: PSALSAR framework, text classification, multimodal processing, transformer, linguistic feature.

INTRODUCTION

The rapid growth of digital platforms, social networking services, and mobile technologies has fundamentally transformed how information is created, distributed, and consumed. In this increasingly interconnected digital environment, news and online content can now be disseminated almost instantaneously across multiple platforms, thereby accelerating the circulation of information among users. Within this evolving landscape, clickbait has become a persistent feature of online media, largely driven by the intense competition for user attention. In general, clickbait refers to content that uses sensational or misleading headlines to provoke curiosity and encourage clicks. The main goal is often to increase engagement rather than provide meaningful information (Raj et al., 2024). While these strategies can increase visibility, they also raise concerns about content quality and credibility.

Clickbait practices are closely linked to platform-driven digital advertising ecosystems, where content visibility and user engagement determine content success. Prior studies have indicated that emotionally appealing and sensationalised content tends to generate higher levels of audience engagement, including clicks, shares, and user interactions (Khawar & Boukes, 2025). In such environments, content performance is frequently evaluated through measurable indicators, such as reach and engagement, which subsequently shape content design strategies. Empirical evidence from digital advertising research further demonstrates how different forms of content are systematically optimised to achieve specific performance objectives, reinforcing the role of engagement-driven mechanisms in shaping online content production (Zahirah et al., 2026). As a result, content creators are often incentivised to prioritise attention-grabbing techniques over informational accuracy.

Other than that, recent studies also indicate that algorithmic distribution systems and digital advertising infrastructures may unintentionally facilitate the dissemination of misleading or low-quality content. Content that generates high engagement rates often receives greater financial rewards and wider exposure (W. Ahmad et al., 2024). Over time, this creates an environment where clickbait practices can continue to grow. It also contributes to the broader circulation of misinformation and manipulative online media behaviour (Hu et al., 2025). In highly competitive digital news environments, content that attracts more interaction tends to gain wider exposure, further reinforcing these patterns (Simonov et al., 2025). Beyond engagement metrics, exaggerated or emotionally charged language can distort expectations, contribute to misinformation, and erode trust in digital journalism (Diez-Gracia et al., 2024; Jacobo-Morales & Marino-Jiménez, 2024). These challenges have driven the development of computational approaches to detect clickbait content.

Existing studies have employed a wide range of techniques, including Natural Language Processing (NLP) (Bajaj & Vishwakarma, 2024), classical machine learning (I. Ahmad et al., 2020; Bronakowski et al., 2023; Daoud & Abou El-Seoud, 2019), and deep learning models (Ghosh, 2021; Naeem et al., 2020; Panda et al., 2024). Beyond clickbait detection, prior research has also demonstrated the effectiveness of analysing user-generated textual data, including social media comments, for extracting meaningful information and improving content understanding (Thabet et al., 2021). This reinforces the broader role of text-based analysis in interpreting digital content and highlights its relevance for tasks such as clickbait detection, where linguistic cues play a central role.

Despite these advances, the existing literature remains methodologically fragmented. Variations in dataset construction, preprocessing techniques, feature representation, and model design have led to inconsistencies in reported findings. Consequently, this lack of methodological consistency makes it

difficult to identify stable methodological patterns or establish reliable benchmarks for cross-study comparisons. This lack of coherence limits the comparability of results and challenges the development of more robust, generalisable detection systems.

To address these issues, this study presents a systematic literature review of computational approaches to clickbait detection reported in studies published between 2016 and 2025. The review draws on literature retrieved from major academic databases, including Scopus, IEEE Xplore, Web of Science (WoS), and SpringerLink. Using a structured review protocol, this study analyses datasets, preprocessing strategies, modelling techniques, linguistic features, and evaluation practices in existing research. The objective is to identify methodological trends, highlight existing limitations, and provide a clearer foundation for future research in developing more effective and generalisable clickbait detection systems.

RELATED WORKS

Existing literature on clickbait detection was reviewed from various perspectives. However, the scope and depth of these reviews have varied considerably. Some studies have emphasised conceptual and behavioural aspects of clickbait. In contrast, others have focused on model performance or specific algorithmic approaches. This variation has resulted in fragmented insights, particularly when attempting to understand methodological developments in a unified manner.

Table 1 highlights differences in publication period, database coverage, number of reviewed articles, and technical focus across previous reviews. Existing reviews differ substantially in both scope and methodological transparency. Several studies provide broad overviews but offer limited technical analysis. In contrast, others focus on performance evaluation without clearly specifying data sources or selection criteria. For example, *Jácono-Morales and Marino-Jiménez (2024)* analysed 165 publications from Scopus and WoS, focusing on conceptual and psychological aspects. While informative, it does not sufficiently address technical design elements such as preprocessing strategies, feature representation, or model architecture. In contrast, *Raj et al. (2024)* examined performance benchmarks across different detection approaches, but lacked transparency in dataset selection and methodology.

Table 1

Comparison of Current Systematic Review with Prior Studies

Study	Period	Database covered	Number of papers	Focus	Key limitation
(<i>Jácono-Morales & Marino-Jiménez, 2024</i>)	2015 -2023	Scopus, WoS	165	Conceptual & Psychological	Limited technical analysis; excludes IEEE & Springer
(<i>Raj et al., 2024</i>)	Not Stated	Not Specified	Not Stated	Performance Benchmarks	Lacks methodological transparency and reproducibility

(continued)

Study	Period	Database covered	Number of papers	Focus	Key limitation
(Yadav & Bansal, 2023)	2016-2022	Not Specified	Not Stated	Machine Learning Overview	Limited analysis of preprocessing and linguistic features
(Zuhroh & Rakhmawati, 2019)	2016-2019	IEEE, ACM, Elsevier, Google Scholar	21	Early Detection Methods	Small sample size; limited temporal coverage
This Study	2016-2025	Scopus, IEEE, WoS, SpringerLink	43	End-to-end Technical Pipeline	-

Yadav and Bansal (2023) provided a more technical perspective on machine learning approaches but offer limited discussion of preprocessing and linguistic features. Similarly, Zuhroh and Rakhmawati (2019) reviewed early studies but are constrained by a small sample size and limited temporal coverage. Overall, existing reviews tend to examine isolated aspects rather than providing an integrated view of end-to-end pipelines. Differences in database coverage and methodological transparency further limit comparability.

METHODOLOGY

This study adopted the Protocol, Search, Appraisal, Synthesis, Analysis, and Report (PSALSAR) framework, which extends the Search, Appraisal, Synthesis, and Analysis (SALSA) approach by incorporating explicit protocol definition and structured reporting to improve transparency and reproducibility (Mengist et al., 2020). It comprises six stages: Protocol, Search, Appraisal, Synthesis, Analysis, and Reporting. The stages ensure the review process remains consistent, traceable, and methodologically sound.

Protocol

The protocol defines the scope and research direction through five Research Questions (RQs):

- RQ1: What datasets are commonly used?
- RQ2: What preprocessing techniques are applied?
- RQ3: What models or algorithms are employed?
- RQ4: What linguistic characteristics are used?
- RQ5: What evaluation metrics are reported?

Peer-reviewed studies were retrieved from four major academic databases: Scopus, IEEE Xplore, WoS, and SpringerLink, selected for their broad coverage of computer science and information technology research.

Search Strategy and Database Selection

The search strategy uses keyword combinations related to “clickbait” and “detection,” including variations such as “click bait” and “click-bait” with detection-related terms (detect, classify, and predict). We applied and adapted Boolean operators for each database. Correspondingly, searches

were conducted across selected databases with filters for publication year and English language. This approach captures both early foundational studies and recent developments, providing a comprehensive starting point for the review process. Queries and results are summarised in Table 2.

Table 2

Search Strings Query

Database	Query	Additional feature	Result
Scopus	TITLE-ABS-KEY ((clickbait OR "click bait" OR Click-Bait) AND (detect* OR identificat* OR estimat* OR classificat* OR predict*))	Year:2016-2025 Language:English	306
IEEE	("clickbait" OR "click bait" OR "Click-Bait") AND ("detect" OR "identific*" OR "estim*" OR "classific*" OR "predict*")	Year:2016-2025 Language:English	161
WoS	TS=(clickbait OR "click bait") AND TS=(detect* OR identificat* OR estimat* OR classificat* OR predict*)	Year:2016-2025 Language:English	157
Springer	(clickbait OR "click bait") AND (detect* OR identificat* OR estimat* OR classificat* OR predict*)	Year:2016-2025 Language:English	22

Study Appraisal and Selection Criteria

The appraisal stage refines the initial dataset by systematically applying relevance and quality criteria, as summarised in Table 3. This stage was conducted in two steps: first, applying predefined inclusion and exclusion criteria; and second, conducting a structured quality assessment to retain only methodologically sound studies. These criteria were established prior to the screening process to promote methodological consistency and reduce potential selection bias.

Table 3

Inclusion and Exclusion Criteria

Type	Criteria
Inclusion	Articles published between 2016–2025. Peer-reviewed journals to guarantee scientific credibility. Publications written in English. Research focusing on clickbait detection, classification, or prediction Research proposing or evaluating datasets, feature extraction methods, or machine/deep learning models, aligning with the research objectives
Exclusion	Studies on clickbait without technical modelling. Editorials, theses, or non-peer-reviewed publications Articles written in languages other than English. Duplicate entries across databases Studies lacking methodological detail, dataset description, or evaluation metrics make replication or comparison impossible.

In this study, the selection emphasises empirical research on clickbait detection from a computational perspective. Only primary research articles published in peer-reviewed journals were included. This decision ensured the studies considered provided sufficient methodological detail. Consistent with the PSALSAR framework (Mengist et al., 2020), non-primary sources such as grey literature, extended abstracts, presentations, and review articles were excluded. These types of publications often lack complete methodological reporting and are therefore less suitable for systematic synthesis. Hence, focusing on fully developed research outputs grounds the analysis in studies that have undergone more extensive scrutiny and refinement.

A similar consideration applies to conference proceedings. While conferences play a central role in disseminating emerging work in computer science and NLP, their format typically prioritises rapid communication over comprehensive reporting. As a result, many conference papers are constrained by strict page limits, which can limit the depth of methodological explanation and experimental transparency (González-Albo & Bordons, 2011). In addition, conference peer review is often shorter and less iterative than journal review, offering fewer opportunities for substantial revision and improvement (Offutt, 2019). This does not imply that conference publications lack value. On the contrary, they are often the first venue for introducing new ideas. Nevertheless, as this review aims to examine methodological trajectories and establish reliable benchmarks, completeness and reproducibility take priority. In this regard, journal articles tend to offer a more stable basis for synthesis, in both methodological transparency and sustained academic impact (Hao et al., 2025).

Quality Assessment

Each study was evaluated using four criteria: scope clarity, completeness of search coverage, peer-review status, and methodological detail, as summarised in Table 4. This step ensures that only studies with sufficient methodological detail and reliability are included, thereby strengthening the validity of the review findings.

Table 4

Quality Assessment Criteria

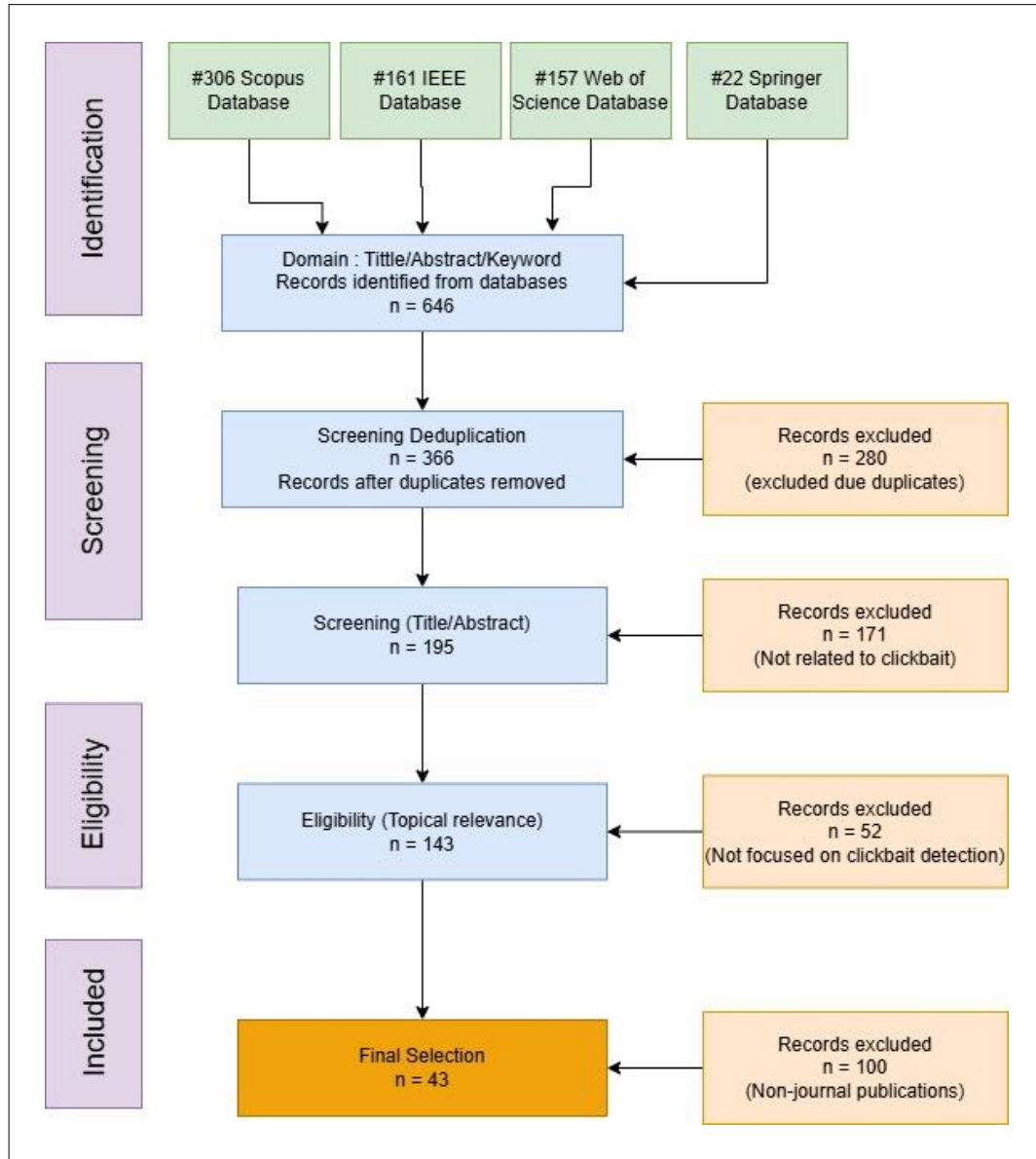
Code	Quality assessment question	Evaluation
QA1	Are the inclusion and exclusion criteria clearly defined?	The study explains how articles were selected and excluded.
QA2	Is the literature search sufficiently comprehensive?	The search covers multiple relevant databases and sources.
QA3	Were the included studies peer-reviewed?	The study was published in a peer-reviewed venue.
QA4	Is the methodology described in sufficient detail?	Study provides sufficient methodological detail, including dataset description, preprocessing steps, model configuration, and evaluation metrics.

Figure 1 illustrates the study selection process. The study initially identified 646 records from multiple academic databases. After removing duplicate entries, 366 unique records remained. Preliminary screening based on titles, abstracts, and keywords excluded 171 records for irrelevance to clickbait. The remaining 195 studies were further assessed for topical relevance, excluding 52 studies

that did not specifically address clickbait detection. Finally, 143 studies were evaluated based on publication type, and only peer-reviewed journal articles were retained, yielding a final corpus of 43 studies for in-depth analysis.

Figure 1

Flow Diagram for the Database Search



Synthesis

The synthesis stage systematically extracted and organised key information using a structured data extraction form. Extracted attributes include bibliographic information, datasets, preprocessing techniques, model approaches, linguistic features, and evaluation metrics, as summarised in Table 5. Data extraction was conducted iteratively. The first pass captured general characteristics, while the second focused on methodological details. Subsequently, studies were categorised by modelling approach and dataset use to support comparison.

Table 5

Study Synthesis

Extraction category	Description	Related RQ
Bibliographic Information	Author(s), publication year, and source type (journal or conference)	-
Dataset Type and Source	Dataset name, availability (public/private), and language characteristics	RQ1
Preprocessing Techniques	Preprocessing methods are applied before modelling	RQ2
Model Approach	Machine learning, deep learning, hybrid, or fusion-based approaches	RQ3
Feature	Linguistic features used in Clickbait detection	RQ4
Evaluation Metrics	Metrics used for model performance evaluation	RQ5
Findings, Challenges, and Opportunities	Main findings, limitations, and potential research gaps identified by the authors	All RQs

Analysis

The analysis stage interprets extracted data through two complementary approaches: narrative synthesis and comparative evaluation. Narrative synthesis identifies recurring patterns in datasets, preprocessing, models, linguistic features, and metrics. Meanwhile, comparative evaluation complements this process by examining similarities and differences across studies. This comparative analysis includes evaluating model performance on similar datasets, identifying commonly used preprocessing pipelines, and examining trends in model complexity and evaluation practices. This combined approach enables a more balanced interpretation, linking qualitative insights with observable methodological patterns.

Reporting

The reporting stage presents the methodology and findings transparently and systematically, documenting the protocol, search process, screening results, and quality assessment. Results are organised by research question and supported by structured tables, enabling reproducibility and clarity.

EVALUATION AND RESULTS

This section presents the findings derived from the selected studies, focusing on dataset characteristics, preprocessing strategies, modelling approaches, linguistic features, and evaluation practices. Rather than listing individual studies, the discussion highlights recurring patterns and methodological tendencies observed across the literature.

Overview of Selected Studies

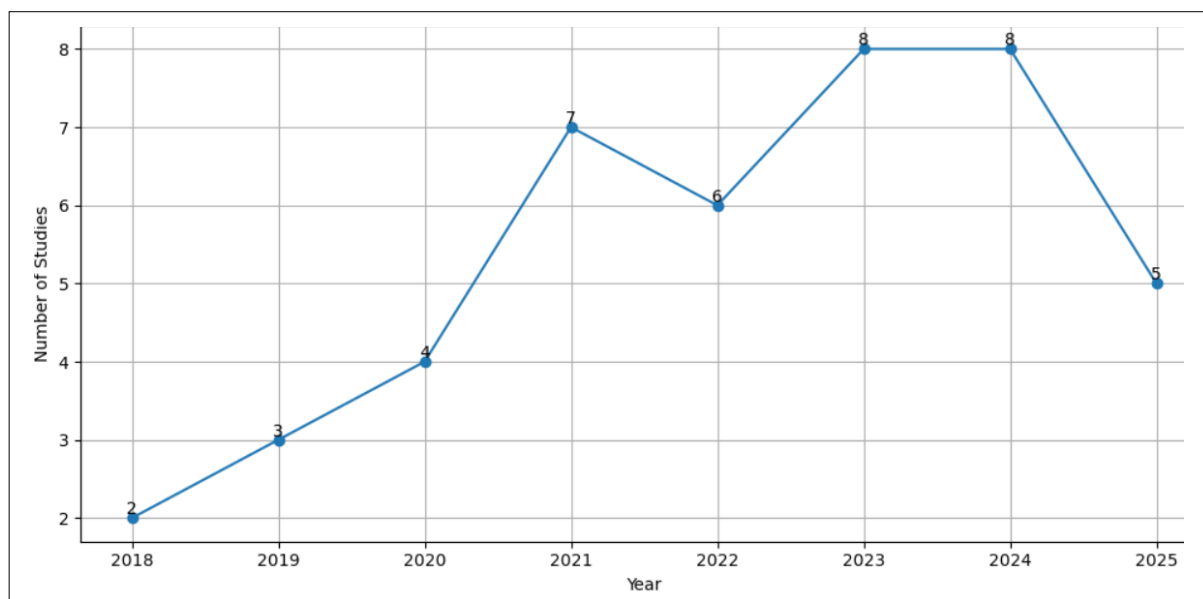
Research on computational clickbait detection began to gain momentum around 2016, a period largely characterised by conference-driven contributions. Early works, such as the dataset introduced by

Chakraborty et al., (2016) in 2016 and the Webis Clickbait Challenge by Potthast et al. (2016, 2018), played a foundational role in shaping the field by defining benchmark datasets and shared evaluation tasks. These studies provide historical and contextual grounding for clickbait detection research.

These conference publications are not part of the systematically selected corpus analysed in this review. Instead, they illustrate the field's early stage, when researchers primarily disseminated outputs through conferences before transitioning to more mature journal publications. Following this initial phase, journal publications began to emerge more consistently after 2018, indicating a shift toward more comprehensive experimentation and detailed methodological reporting. Figure 2 illustrates this transition by reflecting the evolution of publication patterns in the field.

Figure 2

Number of Clickbait Distribution Studies each Year



In line with the selection criteria defined in this study, only peer-reviewed journal articles were included in the final analysis. The absence of journal publications before 2018 is not due to omission. However, it reflects the field's publication dynamics, where early research contributions were predominantly presented at conferences before being extended into journal articles.

RQ1: Dataset

Dataset analysis reveals a strong preference for textual data, particularly short-form content such as news headlines and social media posts. As summarised in Table 6, only a small number incorporate additional signals such as metadata or user interactions, indicating that clickbait detection remains largely text-centric.

Table 6

Clickbait Dataset

No.	Studies	Language	Type	Class and size of dataset	Public availability
P1	(H.-T. Zheng et al., 2018)	Chinese	Text	Clickbait:7.461 Non-Clickbait:7.461	√
P2	(López-Sánchez et al., 2018a)	English	Text	Clickbait:16.000 Non-Clickbait: 16.000	√
P3	(Daoud & Abou El-Seoud, 2019)	English	Text	Clickbait:15.536 Non-ClickBait:6.497	√
P4	(Shang et al., 2019)	English	Text comment and metadata video	Clickbait:141 Non-ClickBait:484	×
P5	(Singh Sisodia, 2019)	English	Text	Clickbait:60.000 Non-ClickBait:40.000	×
P6	(I. Ahmad et al., 2020)	English	Text	Clickbait:16.000 Non-Clickbait:16.000	×
P7	(Naeem et al., 2020)	English	Text	Clickbait:16.000 Non-Clickbait: 16.000	√
P8	(Kaur et al., 2020)	English	Text	Dataset 1 Clickbait: 15.999 Non-Clickbait:16.001 Dataset 2: Clickbait: 5.637 Non-Clickbait:6.080 Dataset 3: Clickbait: 1.200 Non-Clickbait:6.00	√
P9	(William & Sari, 2020)	Indonesian	Text	Clickbait:6.290 Non-ClickBait:8.710	√
P10	(Rajapaksha et al., 2021)	English	Text	Dataset 1:Webis-Clickbait-17 (Twitter posts) Dataset 2: Kaggle clickbait news dataset In both datasets, clickbait is a minority class: 1:3 vs. non-clickbait in Webis-Clickbait-17 and 1:4 in the Kaggle set.	√
P11	(Pujahari & Sisodia, 2021)	English	Text	Clickbait: 16.000 Non-Clickbait: 16.000	√
P12	(Varshney & Vishwakarma, 2021)	English	Video, comment, channel profile	Clickbait: 474 Non-Clickbait: 513	×

(continued)

No.	Studies	Language	Type	Class and size of dataset	Public availability
P13	(J. Zheng et al., 2021)	Chinese & English	Text	Dataset 1 Chinese Clickbait: 1.500 Non-Clickbait:3.384 Dataset 2 English: Clickbait: 4761 Non-Clickbait:14.777	×
P14	(Mundotiya & Yadav, 2021)	English	Text	Clickbait: 5.523 Non-Clickbait: 16.510	√
P15	(Al-Sarem et al., 2021)	Arabic	Text	Clickbait: 23.981 Non-Clickbait: 30.912	√
P16	(Ghosh, 2021)	English	Text	Clickbait: 5.523 Non-Clickbait: 16.510	√
P17	(Razaque, Alotaibi, Alotaibi, Amsaad, et al., 2022)	English	url links	-	×
P18	(Wei & Nguyen, 2022)	English	Text	Clickbait: 5.523 Non-Clickbait: 16.474	√
P19	(Razaque, Alotaibi, Alotaibi, Hussain, et al., 2022)	English	Text; URL Links	Dataset 1: Clickbait: 16.000 Non-Clickbait: 16.000 Dataset 2: 38.517 Twitter Post	√
P20	(Zhou et al., 2022)	Chinese	Text	Clickbait: 4.139 Non-Clickbait: 4.037	×
P21	(Liu et al., 2022)	Chinese	Text	3 Class: General Clickbait: 4.854 Non-Clickbait: 2.427 Malicious Clickbait:2.427	√
P22	(Fakhruzzaman & Gunawan, 2022)	Indonesian	Text	Clickbait: 3.000 Non-Clickbait: 3.000	√
P23	(Meng et al., 2023)	English	Text	Dataset 1: Clickbait: 4.760 Non-Clickbait: 14.730 Dataset 2: Clickbait: 2.620 Non-Clickbait: 3.132	×
P24	(Genç & Surer, 2023)	Turkish	Text	Clickbait: 24.031 Non-Clickbait: 24.029	√
P25	(Andono et al., 2023)	Indonesian	Text	3 Class: Clickbait: 3.316 Non-Clickbait: 5.297 Gray Clickbait: 6.387	√
P26	(Bronakowski et al., 2023)	English	Text	Clickbait: 16.000 Non-Clickbait: 16.000	√
P27	(Supriya et al., 2023)	English	Text	Dataset 1: Clickbait: 16.000 Non-Clickbait: 16.000 Dataset 2: 22.033 Twitter Post	√

(continued)

No.	Studies	Language	Type	Class and size of dataset	Public availability
P28	(Natanya & Liebeskind, 2023)	Hebrew	Text	Clickbait: 700 Non-Clickbait: 700	√
P29	(Fakhruzzaman et al., 2023)	Indonesian	Text	Clickbait: 3.316 Non-Clickbait: 3.316	√
P30	(Chowanda et al., 2023)	English	Text	Clickbait: 16.000 Non-Clickbait: 16.000	×
P31	(Alhanaya et al., 2024)	Arabic	Text	Clickbait: 7.735 Non-Clickbait: 2.829	√
P32	(Ma et al., 2024)	Chinese	Text	Clickbait: 6.000 Non-Clickbait: 6.000	×
P33	(H.-C. Wang et al., 2024)	Chinese	Text	Clickbait: 2.919 Non-Clickbait: 17.977	×
P34	(Imran et al., 2024)	Bangla	Text	Clickbait: 29.312 Non-Clickbait: 223.758	√
P35	(Panda et al., 2024)	English	Text	Clickbait: 2.000 Non-Clickbait: 2.000	√
P36	(Bajaj & Vishwakarma, 2024)	English	Text	Clickbait: 16.000 Non-Clickbait: 16.000	√
P37	(R. R. Sharma et al., 2024)	Amharic	Text	Clickbait: 26.678 Non-Clickbait: 26.549	√
P38	(C. Sharma et al., 2024)	English	Video Metadata (Text, Numeric metadata, Images)	25.856 Video Class imbalance ratio of 2.1 : 1 (non-Clickbait: clickbait) Dataset 1 :Clickbait: 16.000 Non-Clickbait: 16.000	√
P39	(Y. Wang, Zhu, Li, Qiang, et al., 2025)	English	Text	Dataset 2: Clickbait: 814 Non-Clickbait: 1.574 Dataset 3: Webis Clickbait-23(W23C) Dataset Tencent: Clickbait: 1.563 Non-Clickbait: 1.697 Dataset Zongxiang: Clickbait: 960 Non-Clickbait: 1.168	×
P40	(Y. Wang, Zhu, Li, Wei, et al., 2025)	Chinese	Text	Dataset Wangyi Clickbait: 1.072 Non-Clickbait: 1.150 Dataset Wechat Clickbait: 11.925 Non-Clickbait: 5.731	√
P41	(Muqadas et al., 2025)	Urdu	Text	Clickbait:500 Non-Clickbait:500	×
P42	(J. Wang et al., 2025)	English	Text	TikTok ICME Challenge & MIND Dataset Size not specified-	×
P43	(Lwin et al., 2025)	English	Text	Clickbait:15.999 Non-Clickbait:16.001	√

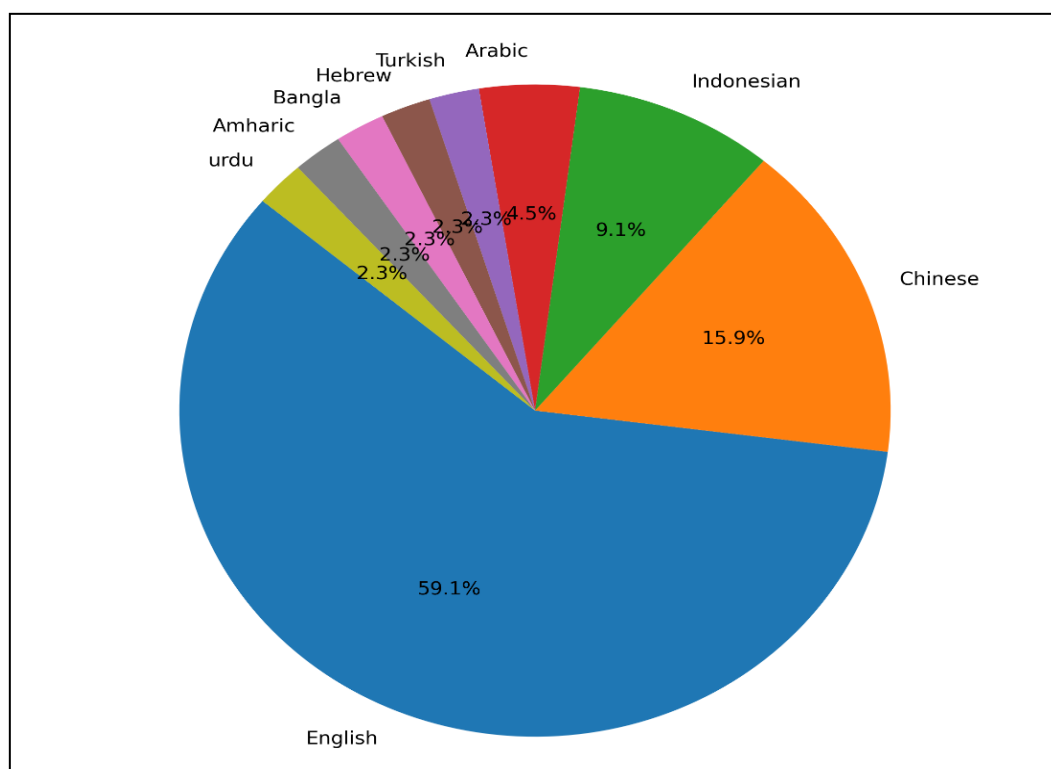
Most studies conceptualise clickbait detection as a binary classification task. This simplification is applied consistently across datasets and languages, indicating a shared assumption that clickbait can be clearly separated from non-clickbait content. Nevertheless, this assumption does not always hold. A small number of studies have introduced a third category to capture ambiguous cases, where headlines display partial clickbait characteristics or receive inconsistent annotations from human evaluators. While these approaches acknowledge the subjective nature of clickbait, they remain limited in adoption. Thus, most existing datasets do not adequately represent annotation uncertainty.

Public datasets available through platforms such as GitHub and Kaggle improve research reproducibility, whereas privately collected datasets may limit cross-study comparability. As Figure 3 illustrates, the distribution of languages across the reviewed datasets is notably imbalanced. English dominates the landscape, followed by Chinese, while other languages appear infrequently. This concentration reflects resource availability rather than the absence of clickbait in other linguistic contexts. For English, multiple independently developed datasets exist, enabling repeated benchmarking and cross-study validation. Chinese datasets indicate a similar, though smaller-scale, pattern.

In contrast, studies involving Indonesian and other low-resource languages tend to rely on a single dataset reused across multiple studies, creating a structural limitation. Performance results in these settings are often dataset-dependent and may not generalise across domains or platforms. Similar patterns have been observed in Arabic, Turkish, and Urdu. Overall, the findings suggest that dataset design in clickbait detection research prioritises availability and convenience over diversity. This has enabled rapid methodological development, but it also constrains evaluations of model robustness across languages and domains.

Figure 3

Clickbait Language Distribution



RQ2: Preprocessing

Preprocessing practices vary widely across studies, both in technique and reporting detail, as illustrated in Table 7. Although common steps such as tokenisation, stop-word removal, and normalisation are frequently mentioned, studies rarely specify how they implement them.

Table 7

Preprocessing of Clickbait Detection

Studies no	Dataset language	Preprocessing activities																Additional information
		T C	L C	R P	S W	L Z	S M	T Z	R S	T N	S G	P S	R N	P F	G R	I N		
P1	Chinese	-	-	-	√	-	-	-	-	-	√	-	-	√	-	-	-	-
P2	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P3	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P4	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P5	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P6	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P7	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P8	English	-	√	-	-	-	-	-	√	√	-	-	-	-	-	-	√	Textual and non-textual input
P9	Indonesian	√	-	√	-	-	-	-	√	-	-	-	-	-	-	-	-	-
P10	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P11	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P12	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Three sets of features are extracted from the video, comments, and channel information. Google Speech API converts speech to text. (Low-level preprocessing steps are not explicitly described)
P13	Chinese & English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P14	English	√	-	√	-	-	-	-	√	-	-	-	-	-	-	-	-	nlTK was used for preprocessing.
P15	Arabic	√	-	√	√	-	√	√	√	√	-	-	-	-	-	-	-	nlTK 3.6.2 was used for preprocessing Arabic text

(continued)

Studies no	Dataset language	Preprocessing activities																Additional information
		T C	L C	R P	S W	L Z	S M	T Z	R S	T N	S G	P S	R N	P F	G R	I N		
P16	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described	
P17	English	-	-	√	√	√	-	-	-	-	-	-	-	-	-	-	-	
P18	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described . Stanford CoreNLP is used for preprocessing the datasets.	
P19	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described	
P20	Chinese	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described	
P21	Chinese	-	-	-	-	-	-	-	-	-	√	√	-	-	-	-	-	
P22	Indonesian	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described	
P23	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described	
P24	Turkish	-	-	-	-	√	√	-	-	-	-	-	-	-	-	-	‘kelime-bol’ library of the Kalbur Project was used for preprocessing	
P25	Indonesian	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described	
P26	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described	
P27	English	√	-	√	-	-	-	-	√	-	-	-	√	-	-	-	-	
P28	Hebrew	-	-	-	√	-	√	-	-	-	-	-	-	-	-	-	Four groups use as features: statistical features based on frequency, structural features on the headline, semantic features, and structural features on the URL	
P29	Indonesian	-	-	-	√	-	√	√	-	-	-	-	-	-	-	-	Stop words removed using an open-source Indonesian stop word remover PySastrawi	
P30	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described	
P31	Arabic	√	-	√	√			√	√	√	-		√	-	-	-	Used the Pyarabic library (https://pyarabic.readthedocs.io/) & the Ruqia library (https://github.com/Ruqyai/Ruqia-Library) to perform Arabic-specific preprocessing, including text normalisation, tokenisation, stemming, and stop-word removal.	
P32	Chinese	-	-	√	√	-	-	√	√	-	√	-	-	-	-	-	-	

(continued)

Studies no	Dataset language	Preprocessing activities															Additional information
		T C	L C	R P	S W	L Z	S M	T Z	R S	T N	S G	P S	R N	P F	G R	I N	
P33	Chinese	-	-	-	-	-	-	-	-	-	√	√	-	-	-	-	-
P34	Bangla	√	-	√	-	-	-	-	√	-	-	-	√	-	-	-	-
P35	English	-	-	-	√	-	√	√	-	-	-	-	-	-	-	-	-
P36	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Preprocessing steps are inherited from the original dataset creators, and the article itself does not provide a detailed description or implement additional preprocessing techniques.
P37	Amharic	-	-	√	√	-	-	-	√	√	-	-	-	-	-	-	-
P38	English	√	-	√	√	-	√	√	-	-	-	-	-	-	-	-	Thumbnail scraping and resizing: thumbnails were scraped from YouTube and resized to reduce memory & training time
P39	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P40	Chinese	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P41	Urdu	-	-	√	√	√	√	√	-	-	-	-	-	-	-	-	-
P42	English	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Low-level preprocessing steps are not explicitly described
P43	English	-	√	√	√	√	-	√	-	-	-	-	√	-	-	-	English stop words were removed using the NLTK stopword list, and tokens were lemmatised using the WordNetLemmatizer.

Notes. TC-Text cleaning (HTML or URL removed), LC-Lowercase (casefolding), RP-Remove Punctuation, SW-Remove stop words, LZ-Lemmatisation, SM-Stemming, TZ-Tokenisation, RS-Remove special characters, TN-Text normalisation, SG-Segmentation, PS-Padding the sequence, RN-Remove Numbers, PF-Part of Speech Filtering, GR-Convert to Grayscale, I-N – Image Noise Remover.

A recurring observation is the absence of a clearly defined standard pipeline. While common operations such as punctuation removal, stop-word filtering, tokenisation, lemmatisation, and stemming are frequently mentioned, studies rarely describe their implementation in detail. In many cases, studies state that they applied preprocessing without specifying the sequence, configuration, or parameter choices. This lack of transparency makes it difficult to assess how preprocessing decisions influence model performance.

Several studies report using established NLP tools, including NLTK, Stanford CoreNLP, and language-specific libraries such as PySastrawi and PyArabic. However, these tools do not guarantee reproducibility, as studies often omit exact preprocessing configurations. As a result, similar toolkits may produce different outputs depending on undocumented settings, introducing hidden variability across studies.

Language characteristics further influence preprocessing choices: English and Indonesian typically rely on whitespace tokenisation, whereas Chinese requires segmentation. These differences limit the transferability of preprocessing pipelines. Beyond text, a small subset of studies has incorporated multimodal inputs, including images, video metadata, and audio signals. These approaches introduce distinct preprocessing requirements that extend beyond traditional text processing. For instance, image-based pipelines may involve resizing or normalisation. Meanwhile, audio-based studies often rely on speech-to-text conversion before textual analysis. Although these methods expand the scope of clickbait detection, they also increase pipeline complexity and reduce comparability with purely text-based approaches.

Overall, the findings indicate that preprocessing remains one of the least standardised components of clickbait detection research. Such variations arise from linguistic differences, dataset characteristics, and inconsistent reporting practices. This lack of standardisation limits reproducibility and complicates cross-study comparison, particularly as observed differences in model performance may stem from modelling techniques and from variations in preprocessing choices.

RQ3: Model and Algorithm

The reviewed studies employ a wide range of modelling approaches, spanning classical machine learning, neural architectures, and more recent transformer-based frameworks. Table 7 summarises these approaches, highlighting how studies combine representation methods and classification strategies.

Table 8

Summary of Models, Algorithms, and Representation Methods Used in Clickbait Detection Studies

No.	Year	Language	Embedding representation	Best algorithm	Type
P1	2018	Chinese	Word2vec	Convolutional Neural Network (CNN)	Deep Learning
P2	2018	English	Word2vec	Hybrid Neural-Case Based Reasoning	Machine Learning
P3	2019	English	Bag of words	SVM	Machine Learning
P4	2019	English	Doc2vec	Adaboost	Machine Learning
P5	2019	English	Handcrafted linguistic features	Ensemble Learning (bagging classifier, random forest classifier, adaboost classifier)	Machine Learning
P6	2020	English	BERT	BERT	Deep Learning
P7	2020	English	Word2vec	LSTM	Deep Learning
P8	2020	English	Glove	Hybrid CNN-LSTM	Deep Learning
P9	2020	Indonesian	-	BiLSTM	Deep Learning
P10	2021	English	Robustly optimised BERT approach (RoBERTa)		Deep Learning
P11	2021	English	Word2vec	SVM	Machine Learning
P12	2021	English	Handcrafted feature vector	J48 Algorithm Based on the Decision Tree Algorithm	Machine Learning

(continued)

No.	Year	Language	Embedding representation	Best algorithm	Type
P13	2021	Chinese & English	Glove	BiLSTM+Attention (using cosine distance to calculate the similarity between headlines and content for the final score)	Deep Learning
P14	2021	English	Sentence embeddings (BERT/ELMo) and word embeddings (GloVe)	LSTM-based deep neural network with dense and sigmoid layers	Deep Learning
P15	2021	Arabic	TF-IDF (Using FV-ANOVA as a feature selection method)	SVM	Machine Learning
P16	2021	English	Glove	Logistic Regression	Machine Learning
P17	2022	English	-	Deep Recurrent Neural Network BiGRU-based sequence encoder. (comprises two parts: (a) human semantic knowledge extraction, and (b) attention mechanisms guided by knowledge graphs)	Deep Learning
P18	2022	English	Glove		Deep Learning
P19	2022	English	Content to vector	Blockchain-enabled deep recurrent neural network	Deep Learning
P20	2022	Chinese		Graph Convolutional Network	Deep Learning
P21	2022	Chinese	BERT +GAT	BiLSTM	Deep Learning
P22	2022	Indonesian	BERT	BERT	Deep Learning
			BERT: word embedding GLU: Title sentence feature extraction		
P23	2023	English	Attention-Fused Deep Relevancy Matching Network (ARMN): Similarity content and body	Neural Network with Softmax (fusion of information from both titles and body contents)	Deep Learning
P24	2023	Turkish	-	BiLSTM	Deep Learning
P25	2023	Indonesian	BERT	Neural network with Focal loss	Deep Learning
P26	2023	English	-	SVM & Gradient-Boosted Decision Trees	Machine Learning
P27	2023	English	Sentence BERT	SVM	Machine Learning
P28	2023	Hebrew	Bag Of Words	SVM	Machine Learning
P29	2023	Indonesian	BERT	Neural Network with Sigmoid Classifier	Deep Learning
P30	2023	English	BERT	BERT	Deep Learning

(continued)

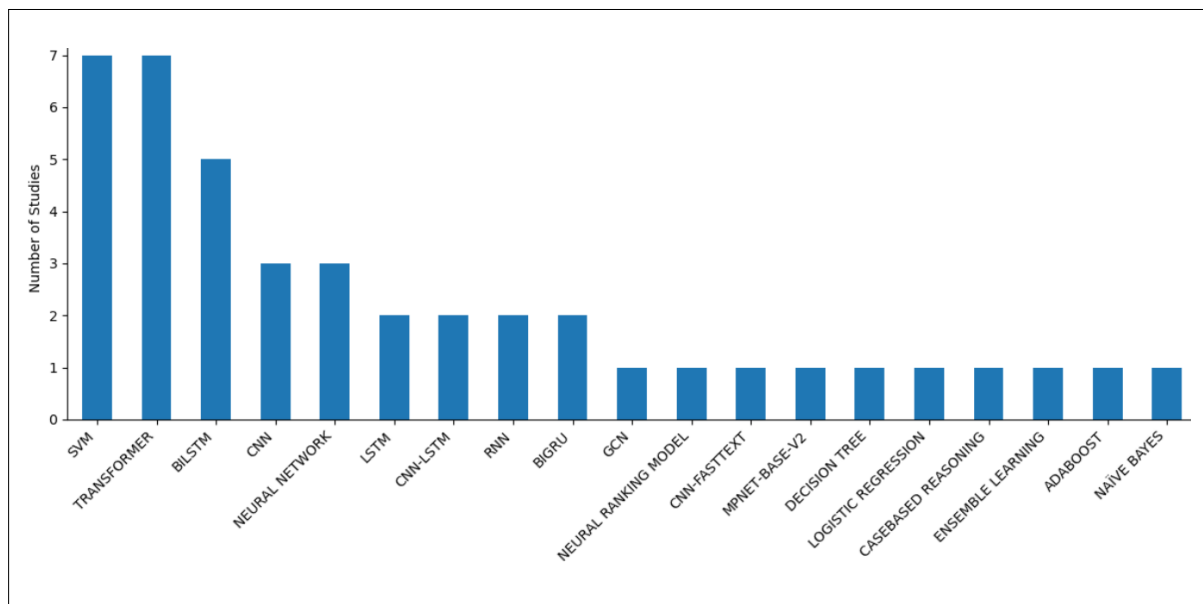
No.	Year	Language	Embedding representation	Best algorithm	Type
P31	2024	Arabic	Word2vec	CNN with AdamOptimizer	Deep Learning
P32	2024	Chinese	Word embedding (Not specifically mentioned)	Hybrid CNN-LSTM	Deep Learning
P33	2024	Chinese	BERT	Bidirectional GRU	Deep Learning
P34	2024	Bangla	-	mpnet-base-v2	Deep Learning
P35	2024	English	BERT	SVM	Machine Learning
P36	2024	English	BERT	RoBERTa	Deep Learning
P37	2024	Amharic	Fasttext	CNN	Deep Learning
P38	2024	English	Used a custom end-to-end deep neural network combining CNN (for images) and FastText (for text+numbers) to perform final binary classification.		Deep Learning
P39	2025	English	Used internal token representations of xlm-roberta-large and BERT within prompt-tuning and verbaliser construction		Deep Learning
P40	2025	Chinese	• XLM-RoBERTa-large as pre-trained language model • Graph Attention Network used to model syntactic • CLIP: Used as an image encoder to get high-level visual features from accompanying images; • Multi-Layer Perceptron is used as a final classifier to output a clickbait vs. non-clickbait label		Deep Learning
P41	2025	Urdu	Sentence Embedding	BiLSTM	Deep Learning
P42	2025	English	Weighted Matrix Factorisation (WMF) for embeddings	Cross Pairwise Ranking (CPR) loss function as prediction layer (combines WMF-based matrix factorisation, dynamic sampling, CPR unbiased ranking loss, and cosine-similarity de-duplication into one integrated framework)	Deep Learning
P43	2025	English	Count Vectoriser-based n-gram	Naïve Bayes and Logistic regression	Machine Learning

A consistent pattern is evident in the use of transformer-based models. In several studies, models such as Bidirectional Encoder Representations from Transformers (BERT) are not used as end-to-end classifiers, but as feature extractors to generate contextual embeddings. Researchers then pass these representations to downstream models, including Bidirectional Long Short-Term Memory (BiLSTM), Gated Recurrent Units (GRU), or Support Vector Machines (SVM), for final classification. To avoid conflating these roles, this review distinguishes between transformers used as primary classifiers and those used purely for representation. This distinction matters as it highlights a broader shift in modelling practice. Rather than replacing earlier architectures entirely, researchers often integrate transformers into existing pipelines as a representational layer.

The distribution of primary classifiers in Figure 4 indicates that both SVM and transformer-based models remain prominent. Their comparable frequency suggests that the field has not fully converged on a single dominant approach. Instead, two paradigms continue to coexist. On the one hand, classical models such as SVMs rely on clearly defined decision boundaries and benefit from well-engineered features. On the other hand, transformer-based models leverage large-scale pretraining to capture contextual meaning. The continued use of SVM, particularly with embeddings such as BERT or word2vec, suggests that simpler models remain effective when supported by strong representations.

Figure 4

Distribution of Primary Classifiers



Recurrent architectures, especially BiLSTM, also appear frequently. These models typically capture sequential dependencies in headline structures, including patterns associated with forward referencing or incomplete information cues. Convolutional models and hybrid CNN–Long Short-Term Memory (LSTM) architectures also appear, though less consistently. Their role is generally limited to extracting local lexical or syntactic patterns, and their usage appears to decline in more recent studies where contextual modelling becomes more prominent.

Beyond these common approaches, several less frequently adopted methods have also been identified, including graph-based models, ensemble techniques, and probabilistic classifiers. These approaches are rarely adopted at scale and are often tied to specific datasets or experimental setups. Their limited adoption suggests these methods remain exploratory rather than established solutions.

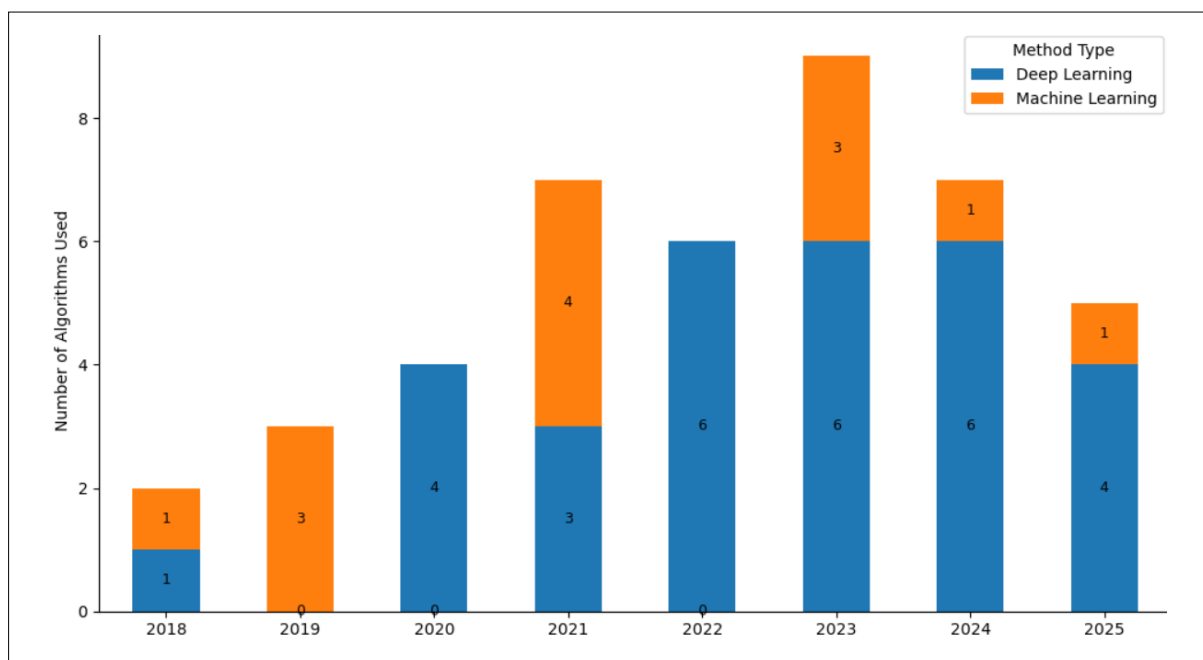
Temporal trends illustrated in Figure 5 indicate a clear transition. Early studies, particularly those published between 2018 and 2019, relied predominantly on classical machine learning supported by handcrafted features. From 2020 onward, deep learning approaches became more prevalent, coinciding with the wider availability of pretrained embeddings and neural sequence models. This shift continues in subsequent years, with deep learning approaches consistently outnumbering classical methods.

More recent studies have introduced prompt-based learning, soft prompt tuning, and multimodal architectures, shifting beyond conventional classification pipelines toward approaches that frame clickbait detection as a task conditioned on instructions or contextual signals. In some cases, textual features are combined with visual or structural information, reflecting a broader interpretation of clickbait as a multimodal phenomenon rather than purely textual manipulation.

Taken together, these findings indicate a gradual but clear evolution in modelling practices. While deep learning and transformer-based approaches have become increasingly dominant in recent work, earlier methods have not disappeared. Instead, they remain competitive when paired with strong feature representations. This suggests that improvements in clickbait detection are driven less by model complexity alone and more by how effectively linguistic and contextual information is represented.

Figure 5

Annual Comparison of Machine Learning and Deep Learning Usage



RQ4: Linguistic Feature

Linguistic features are consistently operationalised as computational inputs rather than descriptive constructs. They influence preprocessing, representation, and model performance. Linguistic features are generally grouped into lexical, syntactic, and semantic levels, following common NLP practice (Nickl et al., 2025; Qiao & Hu, 2019). Building on this foundation, the studies fall into four categories, as illustrated in Table 9: stylistic-lexical, syntactic, semantic (including contextual incongruity), and multimodal linguistic extensions. While the first three categories reflect established feature hierarchies, the multimodal category captures recent developments where textual signals are analysed alongside visual or audio information.

Lexical and stylistic features are among the most commonly employed features in clickbait detection, as they capture attention-grabbing linguistic characteristics such as emotional wording and exaggeration. Some studies explicitly engineer these features, while others rely on deep models to learn them implicitly. Note that syntactic features reflect structural patterns such as interrogatives, imperatives, and incomplete clauses. These features are either explicitly encoded or learned by sequence models and typically complement lexical cues.

Semantic features focus on meaning-level manipulation, particularly phenomena such as the curiosity gap and inconsistencies between headlines and content. Unlike lexical or syntactic features, semantic features often require the joint modelling of multiple inputs, such as headlines, article bodies, or metadata. As a result, their extraction and interpretation are closely associated with representation strategies such as similarity modelling and contextual embedding alignment. Multimodal extensions integrate textual signals with visual or audio cues, reflecting a broader interpretation of clickbait.

Table 9

Linguistic Features Utilised in Clickbait Detection Models

Linguistic aspect	Description	Studies
Stylistic and Lexical Features	Use of attention-grabbing and emotionally charged wording designed to provoke curiosity or exaggeration. For example, sensational words, superlatives, emotional expressions, and excessive punctuation	(Al-Sarem et al., 2021; Bronakowski et al., 2023; Daoud & Abou El-Seoud, 2019; Genç & Surer, 2023; Kaur et al., 2020; Lwin et al., 2025; Ma et al., 2024; Naeem et al., 2020; Natanya & Liebeskind, 2023; R. R. Sharma et al., 2024; Singh Sisodia, 2019; Wei & Nguyen, 2022; J. Zheng et al., 2021)
Syntactic Features	Structural patterns of headlines that deviate from neutral news style to increase engagement. For example, interrogative sentences, imperative forms, short or incomplete clauses, unusual word order	(Al-Sarem et al., 2021; Liu et al., 2022; López-Sánchez et al., 2018b; Muqadas et al., 2025; Naeem et al., 2020; R. R. Sharma et al., 2024; Y. Wang, Zhu, Li, Qiang, et al., 2025; Y. Wang, Zhu, Li, Wei, et al., 2025)
Semantic Features (Including Contextual Incongruity)	Manipulation of meaning through information withholding or mismatch. Relationship between headline and surrounding content or metadata	(Al-Sarem et al., 2021; Ghosh, 2021; Liu et al., 2022; López-Sánchez et al., 2018a; Meng et al., 2023; Mundotiya & Yadav, 2021; Natanya & Liebeskind, 2023; Panda et al., 2024; Supriya et al., 2023; H.-C. Wang et al., 2024; J. Wang et al., 2025; J. Zheng et al., 2021)
Multimodal Linguistic Extensions	Clickbait cues emerge from interactions between textual headlines and other modalities. For example, provocative thumbnails, image-text contradiction, audio-derived cues	(Kaur et al., 2020; Shang et al., 2019; C. Sharma et al., 2024; Varshney & Vishwakarma, 2021; Y. Wang, Zhu, Li, Wei, et al., 2025)

Language differences further influence feature extraction. Languages without explicit word boundaries, such as Chinese, require language-specific segmentation procedures, whereas morphologically rich languages require customised preprocessing. These variations highlight the dependence between linguistic characteristics and preprocessing strategies, reinforcing the connection between RQ4 and RQ2.

RQ5: Evaluation Metrics

Evaluation is primarily framed as a classification problem. As Table 10 illustrates, most studies rely on accuracy, precision, recall, and F1-score to assess model performance across different modelling approaches.

Table 10

Evaluation Metrics in Clickbait Detection Models

No	Evaluation metric						Other evaluation
	Accuracy	F-Measure / F1	Recall	Precision	AUC	ROC	
P1	√	-	√	√	-	-	-
P2	√	-	√	√	-	√	-
P3	-	√	√	√	√	√	-
P4	√	√	√	√	-	-	Kappa, Matthews Correlation Coefficient (MCC)
P5	√	√	√	√	-	-	
P6	√	√	√	√	-	-	
P7	√	√	√	√	-	-	
P8	√	√	√	√	√	√	
P9	√	-	-	-	-	-	-
P10	√	√	√	√	-	-	MCC
P11	√	√	√	√	√	√	-
P12	√	√	√	√	√	√	-
P13	√	√	-	-	-	-	-
P14	√	√	√	√	√	√	-
P15	√	√	√	√			-
P16	√	√	√	√	√	√	-
P17	√	-	-	-	-	-	-

(continued)

No	Evaluation metric						Other evaluation
	Accuracy	F-Measure / F1	Recall	Precision	AUC	ROC	
P18	√	√	√	√	-	-	-
P19	√	-	-	-	-	-	-
P20	√	√	√	√	-	-	RMSE, MAE
P21	√	√	√	√	-	-	-
P22	-	-	-	-	√	√	-
P23	-	√	√	√	-	-	-
P24	√	√	√	√	-	-	-
P25	√	-	-	-	-	-	-
P26	√	-	√	√	-	-	-
P27	√	√	√	√	-	-	-
P28	√	√	√	√	-	-	-
P29	√	√	√	√	√	√	-
P30	√	√	√	√	√	-	Classification Error, Sensitivity, Specificity
P31	√	√	√	√	-	-	-
P32	√	√	√	√	-	-	-
P33	√	√	√	√	-	-	-
P34	√	√	-	-	-	-	Kappa
P35	√	√	√	√	-	-	MCC
P36	√	-	-	-	-	-	PostAttack Accuracy, Attack Success Rate, and Average Perturbed Rate.
P37	√	√	√	√	-	-	-
P38	√	√	-	-	-	-	-
P39	√	√	√	√	-	-	-

(continued)

No	Evaluation metric						Other evaluation
	Accuracy	F-Measure / F1	Recall	Precision	AUC	ROC	
P40	✓	✓	✓	✓	-	-	-
P41	✓	✓	✓	✓	✓	✓	-
P42	-	-	-	-	-	-	Recall@K and NDCG@K
P43	✓	✓	✓	✓	✓	✓	-

Accuracy is frequently reported, particularly in studies that employ balanced binary datasets, where clickbait and non-clickbait instances are represented in relatively similar proportions (Table 6). In such cases, accuracy provides a straightforward measure of overall classification correctness. Nevertheless, relying solely on accuracy can be misleading, especially when class distributions are skewed. For this reason, most studies complement accuracy with precision and recall, which highlight the trade-off between false positives and false negatives, and the F1-score, which balances both. This distinction becomes more important in imbalanced settings. In clickbait detection, false-positive and false-negative errors may have different implications: false positives may incorrectly classify legitimate content as clickbait, whereas false negatives may allow misleading headlines to go undetected. As a result, several studies prioritise F1-score as the main evaluation metric, because it reflects performance under uneven class distributions.

Some studies adopt MCC for more balanced evaluation, particularly in small or imbalanced datasets. Although less frequently reported, its use reflects growing attention to evaluation robustness. Threshold-independent metrics such as ROC and AUC assess model performance across decision thresholds. A small number of studies also employ ranking-based metrics such as Recall@K and NDCG@K for recommendation-oriented scenarios. Overall, metric selection depends on dataset characteristics and task formulation, with classification metrics remaining dominant.

Discussions

This review indicates that research on computational clickbait detection has advanced considerably. However, several methodological inconsistencies continue to persist across the literature. The synthesis across datasets, preprocessing strategies, modelling approaches, linguistic features, and evaluation metrics reveals both convergence in modelling trends and fragmentation in implementation practices.

Most studies rely on English datasets, many of which are reused across multiple studies. This practice supports consistent benchmarking and performance comparisons. Nonetheless, it also limits the generalisability of findings across languages and domains. In contrast, low-resource languages such as Indonesian, Arabic, and Bangla appear in few datasets and are often reused within isolated studies. This imbalance suggests that reported performance improvements may reflect familiarity with the dataset rather than true model robustness. Thus, cross-lingual and cross-domain applicability remains an open challenge.

Preprocessing techniques lack standardisation. Steps such as tokenisation, stop-word removal, and normalisation are frequently mentioned. Nevertheless, authors often omit implementation details, reducing reproducibility and limiting cross-study comparison. Preprocessing pipelines also vary according to linguistic characteristics. Languages without explicit word boundaries require segmentation, whereas morphologically rich languages demand customised normalisation. These findings indicate that preprocessing is not merely a preparatory stage, but a component that directly shapes feature representation and downstream modelling performance.

From a modelling perspective, the findings confirm a shift from classical machine learning toward deep learning and transformer-based approaches. Earlier studies relied heavily on manually engineered features combined with algorithms such as SVM. In contrast, recent work increasingly adopts neural architectures that learn representations directly from data. Nonetheless, classical models remain competitive when paired with effective feature representations. Transformer-based models are also frequently used as embedding generators rather than standalone classifiers, indicating the growing adoption of hybrid architectures. Linguistic features continue to play a central role across modelling approaches. Lexical and stylistic cues remain dominant, while syntactic and semantic features provide complementary contextual information. In traditional machine learning approaches, researchers typically engineer these features explicitly. In contrast, deep learning models capture them implicitly through embeddings and sequence modelling.

A notable gap in the literature is the limited integration of explicitly extracted linguistic features with advanced embedding-based models. Studies that focus on linguistic features often employ classical machine learning techniques. Meanwhile, transformer-based approaches tend to rely on implicit feature learning. This separation indicates an opportunity to combine rule-based linguistic features with hybrid embedding strategies to improve both representation quality and interpretability.

Proposed Conceptual Framework for Computational Clickbait Detection

Based on the synthesis of datasets, preprocessing strategies, linguistic representations, modelling approaches, and evaluation practices identified across the reviewed studies, this review proposes a conceptual framework to guide future research on computational clickbait detection. The proposed framework integrates language-aware preprocessing, linguistic representation, hybrid feature fusion, and adaptive modelling strategies to address current limitations related to generalisability, interpretability, multilingual applicability, and multimodal adaptation. Figure 6 illustrates the proposed conceptual framework developed from the findings of this systematic literature review.

The framework conceptualises clickbait detection as an integrated linguistic and contextual modelling task rather than a purely classification-oriented problem. It highlights the interdependence between preprocessing, linguistic representation, feature learning, and evaluation strategies across different language and dataset conditions, particularly in multilingual and low-resource environments.

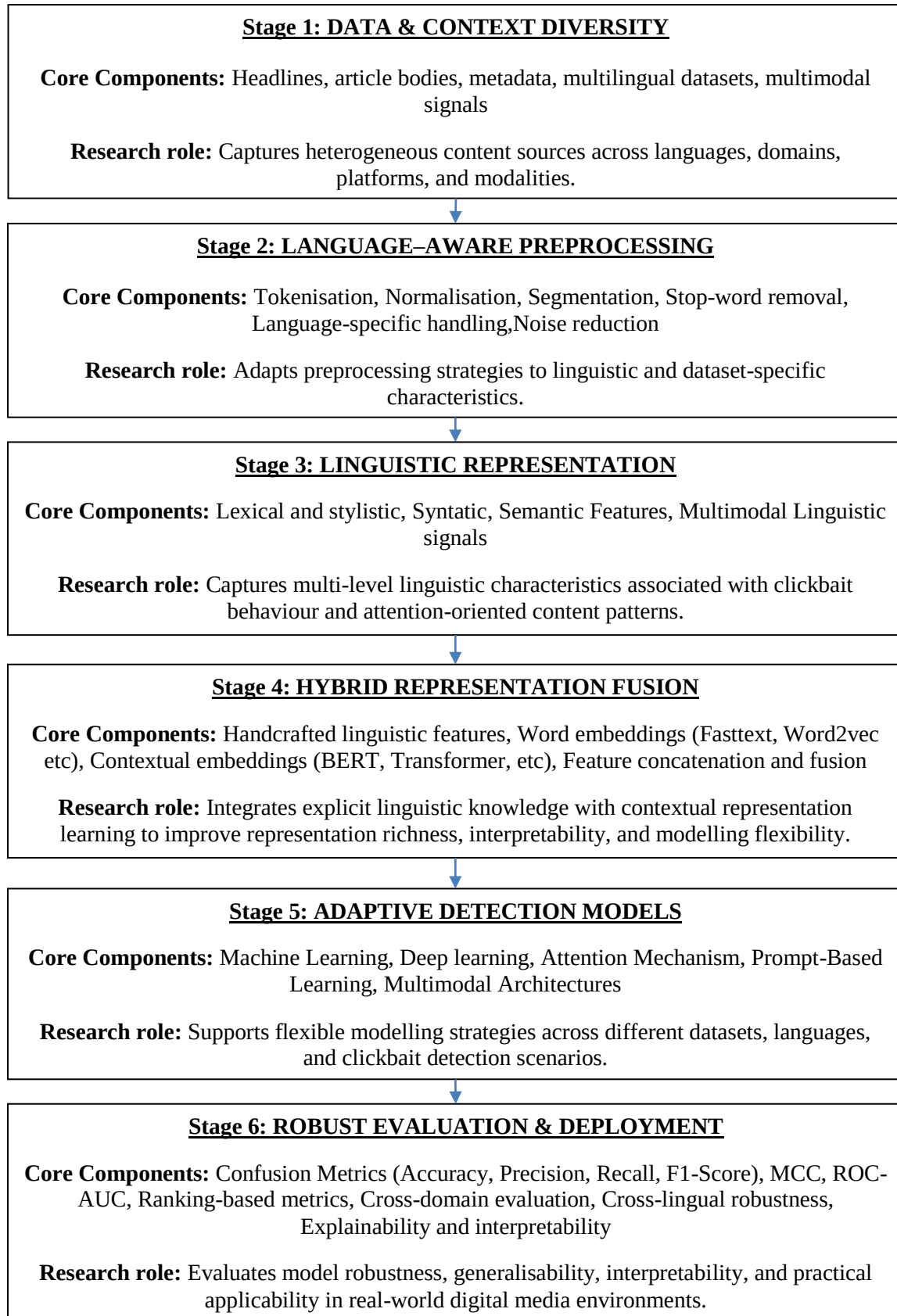
A key contribution of the framework is its hybrid representation-fusion layer, which integrates explicitly engineered linguistic features with contextual embedding approaches, including FastText, Word2Vec, BERT, and transformer-based representations. This integration addresses a methodological gap identified across the reviewed studies, where traditional linguistic features and deep contextual embeddings are commonly implemented separately. The framework, therefore, proposes a more integrative direction for future clickbait detection systems that balances representation richness, interpretability, and modelling flexibility.

The proposed framework is intended as a modular conceptual architecture rather than a fixed implementation pipeline. Not all components need to be used simultaneously. Instead, the framework synthesises the methodological dimensions identified across the reviewed studies and highlights potential integration pathways for future research.

This conceptual framework is the review's primary contribution. Unlike previous reviews that examined datasets, algorithms, or linguistic features separately, this study synthesises these components within a unified methodological perspective. The framework also provides a foundation for future research on hybrid clickbait detection systems that integrate rule-based linguistic features, contextual embeddings, and neural sequence modelling within a single architecture.

Figure 6

Proposed Conceptual Framework for Computational Clickbait Detection Research



Evaluation practices are relatively consistent across studies, with most research relying on conventional classification metrics such as accuracy, precision, recall, and F1-score. While this supports comparability, it also reflects a narrow formulation of clickbait detection as a classification problem. More recent studies have begun to explore ranking-based metrics and multimodal evaluation, but these approaches remain limited.

Overall, the findings suggest that future research should move beyond isolated improvements to individual models or datasets and instead prioritise integrated methodological approaches that combine linguistic insight, robust preprocessing, and advanced representation learning. Hence, such integration is essential for developing clickbait detection systems that are both effective and generalisable in real-world environments.

CONCLUSION

This study presents a systematic literature review of computational approaches to clickbait detection, covering studies published between 2016 and 2025. Using a structured PSALSAR-based methodology, the review examines datasets, preprocessing techniques, modelling approaches, linguistic features, and evaluation metrics to develop an integrated understanding of methodological developments in the field.

The findings demonstrate a clear evolution from feature-based machine learning methods toward deep learning and transformer-based models. Despite this shift, linguistic information remains fundamental, either as implicitly learned representations within neural architectures or as explicit features. The review also identifies several persistent methodological challenges, including limited dataset diversity across languages, inconsistent preprocessing practices, and the dominance of classification-based evaluation frameworks.

A gap remains between linguistic feature engineering and embedding-based modelling. While both approaches have proven effective, researchers rarely integrate them into a unified framework. This suggests a promising direction for future research, particularly in combining rule-based linguistic features with hybrid embedding methods to enhance representation richness and model interpretability.

Future research may extend this work by incorporating conference publications and additional databases to capture emerging developments. More diverse, publicly available datasets are also needed, particularly for underrepresented languages. Thus, incorporating attention-based mechanisms and multimodal information may further improve models' ability to capture complex clickbait patterns, especially for headline-level detection.

ACKNOWLEDGMENT

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- Ahmad, I., Alqarni, M. A., Almazroi, A. A., & Tariq, A. (2020). Experimental evaluation of clickbait detection using machine learning models. *Intelligent Automation & Soft Computing*, 26(6), 1335–1344. <https://doi.org/10.32604/iasc.2020.013861>
- Ahmad, W., Sen, A., Eesley, C., & Brynjolfsson, E. (2024). Companies inadvertently fund online misinformation despite consumer backlash. *Nature*, 630(8015), 123–131. <https://doi.org/10.1038/s41586-024-07404-1>
- Al-Sarem, M., Saeed, F., Al-Mekhlafi, Z. G., Mohammed, B. A., Hadwan, M., Al-Hadhrami, T., Alshammari, M. T., Alreshidi, A., & Alshammari, T. S. (2021). An improved multiple features and machine learning-based approach for detecting clickbait news on social networks. *Applied Sciences*, 11(20), 9487. <https://doi.org/10.3390/app11209487>
- Alhanaya, R., Alqarawi, D., Alharbi, B., & Ibrahim, D. M. (2024). Mushakkal: Detecting Arabic clickbait using CNN with various optimizers. *Journal of Information Technology Management*, 16(4), 64–78. <https://doi.org/10.22059/JITM.2024.99051>
- Andono, P. N., Hadi, P. S., Muljono, M., & Supriyanto, C. (2023). Clickbait detection in Indonesian news title with gray unbalanced class based on BERT. *Journal of Advances in Information Technology*, 14(2), 233–241. <https://doi.org/10.12720/jait.14.2.233-241>
- Bajaj, A., & Vishwakarma, D. K. (2024). Non-alpha-num: A novel architecture for generating adversarial examples for bypassing NLP-based clickbait detection mechanisms. *International Journal of Information Security*, 23(4), 2711–2737. <https://doi.org/10.1007/s10207-024-00861-9>
- Bronakowski, M., Al-khassaweneh, M., & Al Bataineh, A. (2023). Automatic detection of clickbait headlines using semantic analysis and machine learning techniques. *Applied Sciences*, 13(4), 2456. <https://doi.org/10.3390/app13042456>
- Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop clickbait: Detecting and preventing clickbaits in online news media. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 9–16. <https://doi.org/10.1109/ASONAM.2016.7752207>
- Chowanda, A., Nadia, N., & Kolbe, L. M. M. (2023). Identifying clickbait in online news using deep learning. *Bulletin of Electrical Engineering and Informatics*, 12(3), 1755–1761. <https://doi.org/10.11591/eei.v12i3.4444>
- Daoud, D. M., & Abou El-Seoud, S. (2019). An effective approach for clickbait detection based on supervised machine learning technique. *International Journal of Online and Biomedical Engineering (IJOE)*, 15(3), 21–32. <https://doi.org/10.3991/ijoe.v15i03.9843>
- Diez-Gracia, A., Sánchez-García, P., Palau-Sempio, D., & Sánchez-Sobradillo, I. (2024). Clickbait contagion in international quality media: Tabloidisation and information gap to attract audiences. *Social Sciences*, 13(8), 430. <https://doi.org/10.3390/socsci13080430>
- Fakhruzzaman, M. N., & Gunawan, S. W. (2022). CekUmpanKlik: An artificial intelligence-based application to detect Indonesian clickbait. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 11(4), 1232–1238. <https://doi.org/10.11591/ijai.v11.i4.pp1232-1238>
- Fakhruzzaman, M. N., Jannah, S. Z., Ningrum, R. A., & Fahmiyah, I. (2023). Flagging clickbait in Indonesian online news websites using fine-tuned transformers. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(3), 2921–2930. <https://doi.org/10.11591/ijece.v13i3.pp2921-2930>
- Genç, Ş., & Surer, E. (2023). ClickbaitTR: Dataset for clickbait detection from Turkish news sites and social media with a comparative analysis via machine learning algorithms. *Journal of Information Science*, 49(2), 480–499. <https://doi.org/10.1177/01655515211007746>
- Ghosh, S. (2021). Identifying click baits using various machine learning and deep learning techniques. *International Journal of Information Technology*, 13(3), 1235–1242. <https://doi.org/10.1007/s41870-020-00473-1>
- González-Albo, B., & Bordons, M. (2011). Articles vs. proceedings papers: Do they differ in research relevance and impact? A case study in the Library and Information Science field. *Journal of Informetrics*, 5(3), 369–381. <https://doi.org/10.1016/j.joi.2011.01.011>

- Hao, L., Yunwei, C., & Biao, Z. (2025). Study on the differences between journal papers and conference papers in the frontier of basic research: Taking the terahertz field as an example. *20th International Conference on Scientometrics & Informetrics*, 1660–1673. https://doi.org/10.51408/issi2025_042
- Hu, B., Mao, Z., & Zhang, Y. (2025). An overview of fake news detection: From a new perspective. *Fundamental Research*, 5(1), 332–346. <https://doi.org/10.1016/j.fmre.2024.01.017>
- Imran, A. Al, Shovon, M. S. H., & Mridha, M. F. (2024). BaitBuster-Bangla: A comprehensive dataset for clickbait detection in Bangla with multi-feature and multi-modal analysis. *Data in Brief*, 53, 110239. <https://doi.org/10.1016/j.dib.2024.110239>
- Jácono-Morales, D., & Marino-Jiménez, M. (2024). Clickbait: Research, challenges and opportunities: A systematic literature review. *Online Journal of Communication and Media Technologies*, 14(4), e202458. <https://doi.org/10.30935/ojcmr/15267>
- Kaur, S., Kumar, P., & Kumaraguru, P. (2020). Detecting clickbaits using two-phase hybrid CNN-LSTM biterm model. *Expert Systems with Applications*, 151, 113350. <https://doi.org/10.1016/j.eswa.2020.113350>
- Khawar, S., & Boukes, M. (2025). Analyzing sensationalism in news on Twitter (X): Clickbait journalism by legacy vs. online-native outlets and the consequences for user engagement. *Digital Journalism*, 13(8), 1482–1502. <https://doi.org/10.1080/21670811.2024.2394764>
- Liu, T., Yu, K., Wang, L., Zhang, X., Zhou, H., & Wu, X. (2022). Clickbait detection on WeChat: A deep model integrating semantic and syntactic information. *Knowledge-Based Systems*, 245, 108605. <https://doi.org/10.1016/j.knsys.2022.108605>
- López-Sánchez, D., Herrero, J. R., Arrieta, A. G., & Corchado, J. M. (2018). Hybridizing metric learning and case-based reasoning for adaptable clickbait detection. *Applied Intelligence*, 48(9), 2967–2982. <https://doi.org/10.1007/s10489-017-1109-7>
- Lwin, N. O., Jain, R., Dal, R., Yan, H. O., Thaw, K. K., & Naung, S. Y. (2025). Text classification for clickbait detection: A model-driven approach using CountVectorizer and ML classifiers. *Journal of Applied Science and Technology Trends*, 6(1), 43–49. <https://doi.org/10.38094/jastt61237>
- Ma, Y.-W., Chen, J.-L., Chen, L.-D., & Huang, Y.-M. (2024). Intelligent clickbait news detection system based on artificial intelligence and feature engineering. *IEEE Transactions on Engineering Management*, 71, 12509–12518. <https://doi.org/10.1109/TEM.2022.3215709>
- Meng, Q., Liu, B., Sun, X., Yan, H., Liang, C., Cao, J., Lee, R. K.-W., & Bao, X. (2023). Attention-fused deep relevancy matching network for clickbait detection. *IEEE Transactions on Computational Social Systems*, 10(6), 3120–3131. <https://doi.org/10.1109/TCSS.2022.3207479>
- Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*, 7, 100777. <https://doi.org/10.1016/j.mex.2019.100777>
- Mundotiya, R. K., & Yadav, N. (2021). Forward context-aware clickbait tweet identification system. *International Journal of Ambient Computing and Intelligence*, 12(2), 21–32. <https://doi.org/10.4018/IJACI.2021040102>
- Muqadas, A., Khan, H. U., Ramzan, M., Naz, A., Alsahfi, T., & Daud, A. (2025). Deep learning and sentence embeddings for detection of clickbait news from online content. *Scientific Reports*, 15(1), 13251. <https://doi.org/10.1038/s41598-025-97576-1>
- Naeem, B., Khan, A., Beg, M. O., & Mujtaba, H. (2020). A deep learning framework for clickbait detection on social area network using natural language cues. *Journal of Computational Social Science*, 3(1), 231–243. <https://doi.org/10.1007/s42001-020-00063-y>
- Natanya, T., & Liebeskind, C. (2023). Clickbait detection in Hebrew. *Lodz Papers in Pragmatics*, 19(2), 427–446. <https://doi.org/10.1515/lpp-2023-0021>
- Nickl, P., Moussaïd, M., & Lorenz-Spreen, P. (2025). The evolution of online news headlines. *Humanities and Social Sciences Communications*, 12(1), 364. <https://doi.org/10.1057/s41599-025-04514-7>
- Offutt, J. (2019). I love journal papers and you should too. *Software Testing, Verification and Reliability*, 29(3). <https://doi.org/10.1002/stvr.1698>

- Panda, I., Singh, J. P., Pradhan, G., & Kumari, K. (2024). A deep learning framework for clickbait spoiler generation and type identification. *Journal of Computational Social Science*, 7(1), 671–693. <https://doi.org/10.1007/s42001-024-00252-z>
- Potthast, M., Gollub, T., Komlossy, K., Schuster, S., Wiegmann, M., Garces Fernandez, E. P., Hagen, M., & Stein, B. (2018). Crowdsourcing a large corpus of clickbait on Twitter. *COLING 2018: 27th International Conference on Computational Linguistics*, 1498–1507. <https://aclanthology.org/C18-1127/>
- Potthast, M., Köpsel, S., Stein, B., & Hagen, M. (2016). Clickbait detection. *Proceedings of the 7th International Conference on Software and Information Engineering* (pp. 810–817). ACM. https://doi.org/10.1007/978-3-319-30671-1_72
- Pujahari, A., & Sisodia, D. S. (2021). Clickbait detection using multiple categorisation techniques. *Journal of Information Science*, 47(1), 118–128. <https://doi.org/10.1177/0165551519871822>
- Qiao, C., & Hu, X. (2019). Text classification for cognitive domains: A case using lexical, syntactic and semantic features. *Journal of Information Science*, 45(4), 516–528. <https://doi.org/10.1177/0165551518802522>
- Raj, R., Sharma, C., Uttara, R., & Animon, C. R. (2024). A literature review on clickbait detection techniques for social media. *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 1–5. <https://doi.org/10.1109/ICRITO61523.2024.10522359>
- Rajapaksha, P., Farahbakhsh, R., & Crespi, N. (2021). BERT, XLNet or RoBERTa: The best transfer learning model to detect clickbaits. *IEEE Access*, 9, 154704–154716. <https://doi.org/10.1109/ACCESS.2021.3128742>
- Razaque, A., Alotaibi, B., Alotaibi, M., Amsaad, F., Manasov, A., Hariri, S., Yergaliyeva, B. B., & Alotaibi, A. (2022). Blockchain-enabled deep recurrent neural network model for clickbait detection. *IEEE Access*, 10, 3144–3163. <https://doi.org/10.1109/ACCESS.2021.3137078>
- Razaque, A., Alotaibi, B., Alotaibi, M., Hussain, S., Alotaibi, A., & Jotsov, V. (2022). Clickbait detection using deep recurrent neural network. *Applied Sciences*, 12(1), 504. <https://doi.org/10.3390/app12010504>
- Shang, L., Zhang, D. Y., Wang, M., Lai, S., & Wang, D. (2019). Towards reliable online clickbait video detection: A content-agnostic approach. *Knowledge-Based Systems*, 182, 104851. <https://doi.org/10.1016/j.knosys.2019.07.022>
- Sharma, C., Singh, G., Muttum, P. S., & Mahajan, S. (2024). CNN-FastText multi-input (CFMI) neural networks for social media clickbait classification. *Recent Advances in Computer Science and Communications*, 17(6), 13–24. <https://doi.org/10.2174/0126662558283914231221065437>
- Sharma, R. R., Sungheetha, A., Haile, M. A., Kedir, A. H., A, R., & Babu G, C. (2024). Clickbait detection for Amharic language using deep learning techniques. *Journal of Machine and Computing*, 4(3), 603–615. <https://doi.org/10.53759/7669/jmc202404058>
- Simonov, A., Valletti, T., & Veiga, A. (2025). Attention spillovers from news to ads: Evidence from an eye-tracking experiment. *Journal of Marketing Research*, 62(2), 294–315. <https://doi.org/10.1177/00222437241256900>
- Singh Sisodia, D. (2019). Ensemble learning approach for clickbait detection using article headline features. *Informing Science: The International Journal of an Emerging Transdiscipline*, 22, 031–044. <https://doi.org/10.28945/4279>
- Supriya, Singh, J. P., & Kumar, G. (2023). Identification of clickbait news articles using SBERT and correlation matrix. *Social Network Analysis and Mining*, 13(1), 153. <https://doi.org/10.1007/s13278-023-01162-0>
- Thabet, M., Ellouze, M., & Zaied, M. (2021). A new approach for video concept detection based on user comments. *Journal of Information and Communication Technology*, 20(4), 629–649. <https://doi.org/10.32890/jict2021.20.4.7>
- Varshney, D., & Vishwakarma, D. K. (2021). A unified approach for detection of clickbait videos on YouTube using cognitive evidences. *Applied Intelligence*, 51(7), 4214–4235. <https://doi.org/10.1007/s10489-020-02057-9>

- Wang, H.-C., Maslim, M., & Liu, H.-Y. (2024). CA-CD: Context-aware clickbait detection using new Chinese clickbait dataset with transfer learning method. *Data Technologies and Applications*, 58(2), 243–266. <https://doi.org/10.1108/DTA-03-2023-0072>
- Wang, J., Chen, Y., Fu, Y., & Wang, S. (2025). A clickbait identification method for unbiased recommendation. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 49(5), 1125–1136. <https://doi.org/10.1177/18758967251356814>
- Wang, Y., Zhu, Y., Li, Y., Qiang, J., Yuan, Y., & Wu, X. (2025). Clickbait detection via prompt-tuning with titles only. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(1), 695–705. <https://doi.org/10.1109/TETCI.2024.3406418>
- Wang, Y., Zhu, Y., Li, Y., Wei, L., Yuan, Y., & Qiang, J. (2025). Multi-modal soft prompt-tuning for Chinese clickbait detection. *Neurocomputing*, 614, 128829. <https://doi.org/10.1016/j.neucom.2024.128829>
- Wei, F., & Nguyen, U. T. (2022). An attention-based neural network using human semantic knowledge and its application to clickbait detection. *IEEE Open Journal of the Computer Society*, 3, 217–232. <https://doi.org/10.1109/OJCS.2022.3213791>
- William, A., & Sari, Y. (2020). CLICK-ID: A novel dataset for Indonesian clickbait headlines. *Data in Brief*, 32, 106231. <https://doi.org/10.1016/j.dib.2020.106231>
- Yadav, K. K., & Bansal, N. (2023). A comparative study on clickbait detection using machine learning-based methods. *2023 International Conference on Disruptive Technologies (ICDT)*, 661–665. <https://doi.org/10.1109/ICDT57929.2023.10150475>
- Zahirah, K. F., Poernomo, W., Yusuf, E., & Bernardus, D. (2026). Evaluating the effectiveness of digital product advertisement type using machine learning and Shapley additive explanations analysis. *Journal of Information and Communication Technology*, 25(1), 79–106. <https://doi.org/10.32890/jict2026.25.1.5>
- Zheng, H.-T., Chen, J.-Y., Yao, X., Sangaiah, A. K., Jiang, Y., & Zhao, C.-Z. (2018). Clickbait convolutional neural network. *Symmetry*, 10(5), 138. <https://doi.org/10.3390/sym10050138>
- Zheng, J., Yu, K., & Wu, X. (2021). A deep model based on lure and similarity for adaptive clickbait detection. *Knowledge-Based Systems*, 214, 106714. <https://doi.org/10.1016/j.knosys.2020.106714>
- Zhou, M., Xu, W., Zhang, W., & Jiang, Q. (2022). Leverage knowledge graph and GCN for fine-grained-level clickbait detection. *World Wide Web*, 25(3), 1243–1258. <https://doi.org/10.1007/s11280-022-01032-3>
- Zuhroh, N. A., & Rakhmawati, N. A. (2019). Clickbait detection: A literature review of the methods used. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 6(1), 1. <https://doi.org/10.26594/register.v6i1.1561>