Supplementary Materials for "Document Classification Pattern Recognition via Information Fusion: A Systematic Review of Multimodal and Multiview Representation Approaches"

Marcin Michał Mirończuk[a]

[a] *National Information Processing Institute, al. Niepodległości 188b, 00-608, Warsaw, Poland*

Abstract

This supplement accompanies the manuscript on information fusion for document classification. It provides (i) the database search queries, (ii) a link to the full replication materials (extracted data, R scripts, and two technical reports), (iii) extended summaries and structured tables that complement the review (domains, tasks, learning paradigms, data labelling and preprocessing, traditional and deep learning methods, model optimisation, and evaluation/validation), and (iv) a taxonomy of multimodal and multiview works with accompanying figures.

Keywords: Information fusion, Document classification, Multimodal learning, Multiview learning, Systematic review, Meta-analysis, Representation learning

1. Supplementary Information — Preface

This document accompanies the manuscript titled *Document Classification Pattern Recognition via Information Fusion: A Systematic Review of Multimodal and Multiview Representation Approaches*. It consolidates resources for transparency and reproducibility: Section 2 presents the database search queries; Section 3 links to a Zenodo repository containing the dataset extracted for the quantitative meta-analysis, the R scripts, and the technical reports; Sections 4 and 5 provide extended summaries and structured tables across application domains, tasks, learning paradigms, data labelling and preprocessing, traditional and deep learning methods, model optimisation, and evaluation/validation; and Section 6 presents the taxonomy of multimodal and multiview works with the accompanying figure. Section, table, and figure numbering refer to this supplement.

2. Database search queries

We conducted the systematic literature search using the Scopus database. We performed the search in three distinct phases, each using a specialised query designed to identify a particular subset of the literature. The precise queries are detailed below.

[*] Corresponding author

Email address: marcin.mironczuk@opi.org.pl (Marcin Michał Mirończuk)

Query 1: Identifying existing reviews of fusion techniques. This query was designed to find existing survey and review articles that focus on multimodal or multiview learning to ensure that our work builds upon prior syntheses.

```
( TITLE ( {multiview} OR {multi view} OR {multiview} OR {information
   fusion} OR {multi modal} OR {multi-modal} OR {multimodal} ) AND KEY (
   review OR overview OR survey ) ) OR ( ( KEY ( {multiview} OR {multi
   view} OR {multiview} OR {information fusion} OR {multi modal} OR
   {multi-modal} OR {multimodal} ) AND TITLE ( review OR overview OR
   survey ) ) ) AND ( LIMIT-TO ( SUBJAREA , "COMP") ) AND ( LIMIT-TO (
   DOCTYPE , "re") )
```

Query 2: Identifying existing reviews of document classification. This query was used to find survey and review articles on the broader topic of document classification to situate our work in the established literature.

```
(TITLE ( "document classification" OR "document categorization" OR "text
   classification" OR "text categorization" ) AND KEY ( review OR
   overview OR survey ) ) OR ( KEY ( "document classification" OR
   "document categorization" OR "text classification" OR "text
   categorization" ) AND TITLE ( review OR overview OR survey ) ) AND (
   LIMIT-TO ( SUBJAREA , "COMP" ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) OR
   LIMIT-TO ( DOCTYPE , "re" ) OR LIMIT-TO ( DOCTYPE , "sh" ) )
```

Query 3: Identifying primary studies for analysis. This was the main query used to find the primary research articles for our qualitative and quantitative analysis. It was designed to find articles that applied multimodal or multiview techniques to the task of document classification.

```
( TITLE-ABS-KEY ( {multi-view}  OR  {multi view} OR  {multiview} OR
   {information fusion}  OR  {multi modal}  OR  {multi-modal} OR
   {multimodal})  AND  ( ABS ( "document classification"  OR  "document
   categorization"  OR  "text classification"  OR  "text categorization"
   )  OR  KEY ( "document classification"  OR  "document categorization"
   OR  "text classification"  OR  "text categorization" ) ) )  OR  (
   TITLE-ABS-KEY ( "document classification"  OR  "document
   categorization"  OR  "text classification"  OR  "text categorization"
   )  AND  ( ABS ( {multi-view}  OR  {multi view} OR {multiview} OR
   {information fusion}  OR  {multi modal}  OR  {multi-modal}  OR
   {multimodal})  OR  KEY ( {multi-view}  OR  {multi view} OR {multiview}
   OR  {information fusion}  OR  {multi modal}  OR  {multi-modal} OR
   {multimodal} ) ) )  AND  ( LIMIT-TO ( SUBJAREA ,  "COMP" ) )
```

3. Replication materials

3.1. Dataset and code

All data extracted for the quantitative meta-analysis, together with the R scripts used to run the statistical analyses and generate figures, are archived on Zenodo (`https://zenodo.org/records/17141560?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImI0ODcwMDUyLTllNjctNDU1OC1iYjRjLWE0YmQwZGEzNmM2ZSIsImRhdGEiOnt9LCJyYW5kb20iOiI2ZThiZTIzZjZhNGNmNDEyOGI3OTJjNTM2OTQxMTQyYiJ9.Av4nmXokgDBkUdz1QaDHiK-cpqNxmWLLoxJkpevPy3ep14pNgcoG6BXVkMsBgaoK1UU6awch9bcRPnySK1lzDw`). The record contains:

- Bibliography — `bibtex-information-fusion-document-classification.bib` used across the analyses.
- Extracted data — Excel workbooks (`2-articles-information-fusion-summarisation-cleaned.xlsx`) with resource- and study-level data used in the meta-analysis.
- R source — `qualitative-analysis.Rmd` and `quantitative-analysis.Rmd`, which load the bibliography and Excel data to compile two technical reports.
- Technical reports (HTML) — `qualitative-analysis.html` and `quantitative-analysis.html`, with supplementary details on methodology, tables, and figures.

Review-only link. A temporary, non-persistent preview URL may be provided to the editors during peer review; it will be replaced by the DOI in the final version.

4. Information fusion review summarisation

Review works with information fusion are described and discussed in the main article in the Related Works (Analysis of Review Works with Fusion) section.

5. Document classification review summarisation

5.1. Application domain

Table 1 categorises various application domains, highlighting core and associated works that contribute to advancements in each field.

Table 1: The table presents a breakdown of different application domains, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	**Core works**	**Associated works**
Social media	[1, 2, 3, 4, 5, 6, 7, 8, 9]	[10, 11]
Customer reviews	[12, 13, 14, 15, 16, 17, 18, 19]	[20, 10]
Construction budgeting	[21]	
Language-specific classification (e.g. for Arabic, Thai, or Portuguese)	[22, 13, 23, 24, 25, 26]	[27]
Healthcare	[1, 4, 28, 29, 30, 31, 32]	[33, 34, 35, 36]
Legal documents		[37, 38]

Document classification adapts to diverse domains by addressing field-specific challenges. Social media tackles noisy, real-time data [11] and evolving language styles [5, 1] to moderate harmful content, including hate speech and COVID-19 misinformation. Customer reviews employ sentiment analysis [12, 16, 14] to extract detailed insights and fake-review detection [15] to combat deceptive practices. To address linguistic diversity, researchers have focused on non–English languages, which has led to advancements in BERT fine-tuning [22, 23] and the application of classification methods to a wider array of languages [27].

Healthcare confronts sparse data and the challenge of classifying short passages of text with analyses of COVID-19 research [29], clinical text [31], and pharmaceutical sentiment [30]. Work in legal domains addresses lengthy passages of text by utilising domain-specific datasets for tasks like classification and summarisation [37], while work in construction domains supports cost estimation and budgeting processes [21].

5.2. Task

Table 2 categorises various tasks, highlighting the core and associated works that contribute to advancements in each field.

Table 2: The table presents a breakdown of different tasks, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	**Works**	**Associated works**
General document classification	[20, 39, 40, 41, 42, 27, 43, 44, 45, 36, 17, 46, 47, 48, 49, 50]	[37, 51, 10, 52, 23, 53, 54, 55, 56, 25]
Web page classification	[57]	
Sentiment analysis	[16, 13, 14, 58, 18, 19]	[37, 12, 20, 23]
Fake news or review detection	[1, 15, 4]	
Rumor detection	[2]	
Personality prediction or author profiling	[3, 6]	
Authorship attribution	[59]	
Semantic document classification	[60]	
Detecting suspicious emails	[61]	
Cyberbullying detection	[11, 9]	[7]
Text summarisation		[37]

Document classification has evolved from traditional machine learning methods, such as feature selection [49], to advanced architectures like transformers and GNNs. While early work focused on general document categorisation [42], the field has since shifted towards deep learning, with a wide array of architectures like CNNs, RNNs, and transformers now commonly applied to document classification tasks [54]. Simultaneously, recent studies have emphasised domain-specific challenges: sentiment analysis [58, 14] tackles language-specific quirks like sarcasm, and misinformation detection [1, 2] combines linguistic and contextual analysis to counter evolving deception tactics. Beyond these applications, document classification enables user-centric tasks such as personality profiling [3] and authorship attribution [59], leveraging stylistic and syntactic patterns to infer traits. Its versatility is further demonstrated in cybersecurity (detecting malicious content), semantic classification [60], and multimodal web analysis [57]. Post-2020, the field has shifted decisively towards hybrid frameworks and pretrained language models [62, 63], balancing efficiency with nuanced task performance. Researchers now prioritise explainability, multilingual adaptability, and scalability—addressing challenges like data imbalance and computational costs—to meet the demands of real-world deployment in dynamic, multilingual environments.

5.3. Learning paradigms

Table 3 categorises various learning paradigms, highlighting core and associated works that contribute to advancements in each field.

Table 3: The table presents a breakdown of different learning paradigms, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	**Works**	**Associated works**
Supervised learning	[39, 41, 33, 42, 53, 64, 45, 49]	[22, 65, 40, 15, 2, 58, 51, 5, 66, 6, 67, 54, 43, 68, 30, 31, 36, 9, 24, 17, 18, 56, 46, 25, 61, 47, 26, 48, 69, 50]
Semi-supervised learning	[10, 33]	[70, 15, 58, 57, 47]
Self-supervised learning (SSL)		[37, 71]
Transfer learning (multitask learning, domain adaptation)	[63, 3]	[12, 13, 20, 14, 2, 58, 23, 28, 29, 11, 8, 30]
Multiple instance learning	[72]	
Reinforcement learning		[53, 54]
Information fusion learning	[3, 4, 73, 19]	[1, 62, 2, 16, 35, 57, 47]

Supervised learning remains the cornerstone of document classification, with foundational works like [45] and [64] establishing robust frameworks for various document classification tasks. However, its reliance on labelled data has spurred interest in alternatives. Semi-supervised methods, exemplified by [10], address annotation bottlenecks by leveraging unlabelled text through graph-based label propagation, self-training, and cotraining. Self-supervised approaches, such as those in [37], further reduce annotation needs through masked language modelling. Transfer learning bridges gaps in specialised domains, with [23] adapting BERT for Arabic text and [28] tailoring pretrained models for biomedical contexts. Meanwhile, frameworks like multiple instance learning [72] enable weakly supervised classification, in which labels apply to document collections rather than individual instances. Reinforcement learning has been explored in document classification to model sequential decision processes, such as selecting critical tokens or determining when to stop reading a passage of text [54].

Recent advances increasingly prioritise information fusion to enhance model robustness—although challenges persist. Multimodal fusion, for instance, remains underexplored in text-centric applications, as highlighted by [1], which emphasises the importance of integrating text with visuals for COVID-19 fake news detection. Key hurdles include the alignment of heterogeneous data streams (e.g. text with images or metadata) and the mitigation of noise in sparse multiview settings. Ensemble fusion strategies, such as hybrid architectures [73], demonstrate promise by combining diverse model outputs or features. For example, [4] combine features from multiple contextual embeddings (BERT, XLNet, and ELMo) and also explores ensemble methods like a voting classifier to improve misinformation detection.

Critical gaps hinder the broader adoption of fusion techniques. Benchmarks for multimodal text corpora are scarce, and many studies default to simplistic fusion rules (e.g. feature concatenation) rather than context-aware mechanisms. Future work must address these limitations to unlock the full potential of fusion in document classification.

5.4. Data labelling

Table 4 categorises various data labelling techniques, highlighting core and associated works that contribute to advancements in each field.

Table 4: The table presents a breakdown of data labelling techniques, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	Works	Associated works
Text annotation and rationales	[65, 38]	
Label structure (one class, binary, multilabel, hierarchical)	[74, 68]	[22, 62, 39, 10, 75, 33, 42, 23, 29, 53, 67, 30, 45, 46, 47, 69]
Label proportion and count (balanced/imbalanced, few-shot, zero-shot)		[22, 1, 12, 13, 76, 11, 7, 43, 30, 31, 9, 17, 61, 72]

Data labelling practices in document classification emphasise three evolving areas: (1) annotation rationales (e.g. direct studies like [65] and [38], which enhance transparency via human-guided labelling; (2) label structures, in which direct works [74, 68] formalise multilabel taxonomies and explore label correlations, supported by indirect research [42, 33] on overlapping categories; and (3) label proportions, with indirect contributions [31, 30] that address the imbalance. Recent direct efforts prioritise sophisticated frameworks for domain-specific challenges like medical text [28] or low-resource languages [23], which reflects a shift towards adaptable, annotation-efficient systems.

5.5. Data preprocessing

Table 5 categorises various data preprocessing methods, highlighting core and associated works that contribute to advancements in each field.

Table 5: The table presents a breakdown of data preprocessing methods, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	Works	
Tokenisation and tokenisers		[42, 27, 50]
Weighting schemas or text embedding	[75, 71, 55]	[74, 37, 20, 39, 3, 2, 41, 10, 52, 4, 53, 27, 11, 54, 30, 44, 45, 36, 17, 60, 47, 48, 50]
Data augmentation	[76]	

Data preprocessing research emphasises weighting schemas/embeddings, with direct studies like [55] (term weighting), [75] (embeddings), and [71] (traditional word embeddings like Word2Vec). While foundational preprocessing steps like tokenisation are detailed extensively in surveys such as [42, 27], data augmentation is less directly explored in the literature [76]. This skew towards text representation methods highlights their foundational role in natural language processing.

5.6. Traditional machine learning

Table 6 categorises various traditional machine learning approaches, highlighting core and associated works that contribute to advancements in each field.

Table 6: The table presents a breakdown of traditional machine learning approaches, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	**Works**	**Associated works**
Feature selection or projection	[40, 77, 51, 66, 78, 67, 34, 56]	[22, 39, 3, 42, 11, 68, 55, 47, 19, 48, 49, 59]
Algorithms (SVM, naïve Bayes, random forests, etc.)	[64, 7, 45, 17, 49]	[22, 74, 1, 20, 21, 39, 15, 77, 16, 58, 41, 33, 4, 42, 5, 53, 6, 27, 43, 57, 31, 36, 9, 24, 46, 25, 61, 26, 48, 69, 50]

Feature selection/projection evolved from foundational variable identification in early indirect studies (e.g. [59], [49], and surveyed in reviews such as [47]) to automated optimisation in direct works like ([78, 40]), prioritising efficiency in tasks such as medical document classification [28] and hate speech detection [7].

Classic algorithms (SVM, naïve Bayes) remain central [64, 45], with direct studies demonstrating their adaptability in sentiment analysis [58] and fake news detection [1], while iterative refinements focus on computational efficiency [43].

5.7. Deep learning

Table 7 categorises various deep learning solutions, highlighting core and associated works that contribute to advancements in each field.

Table 7: The table presents a breakdown of deep learning solutions, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	**Works**	**Associated works**
Transformer-based models	[63, 2, 37, 62, 12, 29, 54]	[22, 74, 1, 13, 14, 39, 41, 4, 42, 23, 28, 53, 11, 43, 8, 30]
Graph neural networks (GNNs)	[70, 52, 79]	[74, 39, 10, 53]
Recurrent neural networks (RNNs, LSTM, GRU)	[54, 79]	[63, 22, 74, 20, 21, 14, 15, 3, 16, 41, 4, 42, 28, 53, 6, 7, 73, 30, 44, 36]
Convolutional neural networks (CNNs)	[54, 79]	[63, 22, 74, 15, 3, 16, 41, 42, 53, 35, 7, 73, 43, 30, 44, 57, 36]
Multilayer neural networks		[35, 54, 36, 49]

Deep learning in document classification has evolved from simple neural networks like perceptrons [49], through CNN and RNN approaches [36], to advanced architectures like transformers and GNNs, with in-depth studies emerging prominently from 2020 onwards. Early research utilised simpler neural networks, while recent contributions have focused on sophisticated frameworks for tasks such as sentiment analysis [16], fake news detection [1], and biomedical text processing [28]. For instance, studies like [62] (transformers) and [70] (GNNs) illustrate the shift towards complex architectures for nuanced language tasks. This trajectory reflects both methodological innovation and expanding applications across domains.

5.8. Model optimisation

Table 8 categorises various model optimisation techniques, highlighting core and associated works that contribute to advancements in each field.

Table 8: The table presents a breakdown of model optimisation techniques, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	**Works**	**Associated works**
Long document classification	[37]	[62]

In this main category, two articles illustrate different angles on optimising classification models for long-document scenarios. Strategies to improve long-document classification are explored directly in [37], emphasising efficiency and scalability in handling extensive text data. Meanwhile, [62] focuses on document classification using transformer-based models and LLMs, which are relevant to the optimisation of model performance. Together, these works underscore the evolving nature of model optimisation for long-text corpora, stressing the need for robust methods that balance performance with resource constraints.

5.9. Evaluation and validation

Table 9 categorises various evaluation and validation methods, highlighting core and associated works that contribute to advancements in each field.

Table 9: The table presents a breakdown of evaluation and validation methods, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	**Works**	**Associated works**
Explainable AI	[65, 38]	[62]
Robustness and adversarial attacks		[76]
Metrics and validation		[74, 58, 41, 33, 42, 66, 11, 7, 54, 43, 44, 79, 45, 31, 36, 9, 47, 19]
Benchmarks and datasets		[63, 22, 74, 37, 1, 12, 62, 13, 20, 70, 14, 40, 15, 2, 3, 58, 41, 41, 51, 10, 75, 33, 52, 4, 42, 23, 28, 5, 29, 66, 53, 6, 27, 11, 67, 7, 73, 54, 8, 30, 57, 24, 18, 25, 72, 26]
Uncertainty quantification	[35]	

In the *evaluation and validation* category, most works describe evaluation metrics for document classification models (e.g. [45, 54]), the need for standardised datasets (e.g. [22, 37]), and methodologies for the validation of AI systems. A growing subset addresses explainability (e.g. [38, 65]) and uncertainty quantification [35], while niche areas like adversarial robustness [76] highlight challenges in real-world deployment. How explainable frameworks enhance transparency is demonstrated in [38], while [35] systematises uncertainty in deep learning. Indirectly, extensive work refines validation practices: studies on performance metrics [58, 43] and benchmarks [37, 20] dominate, which reflects a shift towards reproducibility. Future efforts must expand standardised test beds and address emerging gaps in fairness and robustness as model complexity grows.

5.10. Challenges and future directions

Table 10 categorises different challenges and future directions, highlighting core and associated works that contribute to advancements in each field.

Table 10: The table presents a breakdown of challenges and future directions, listing significant works categorised as either core or associated contributions. Core works are those focused primarily on these particular topics; associated works supplement the core studies by expanding on related aspects or applications.

Topics	**Works**	**Associated works**
Dealing with multilingual content (cross-lingual and multilingual classification)		[1, 13, 14, 15, 58, 23, 79, 18]
Improving model interpretability	[38, 65]	[54]
Ethical considerations: addressing ethical concerns (bias, fairness)		[62, 8]
Handling of out-of-distribution data		[35]
Prompt engineering	[12]	
Proper set of benchmarks		[74, 39, 10, 11, 67, 43, 72]

Document classification has advanced significantly, driven by methodological innovation and real-world needs. Works exist that address particular non–English languages like Arabic—even challenging their traditional classification as low-resource (e.g. [23, 13]). Recent efforts have focused on ethical considerations, mitigating bias (e.g. [62, 8]), and enhancing model transparency and robustness through interpretability [38] and uncertainty quantification [35]. This push for reliability is supported by the promotion of standardised benchmarks (e.g. [74, 39]). Architectural advancements include a comparison of the performance of fine-tuning and prompt engineering for classification tasks [12] and refined graph-based approaches [70], with direct applications in areas like fake news detection [1] and biomedical text processing [28] emerging since 2020. This trajectory reflects a dual focus: improving foundational aspects like multilinguality and interpretability, while developing architectures for fairness, robustness, and specialised applications.

6. Taxonomy of multimodal and multiview works

Based on the qualitative analysis of the collected works, we developed the taxonomy that is illustrated in Figure 1. The taxonomy classifies the studies according to key characteristics such as modalities, fusion techniques, and evaluation metrics.

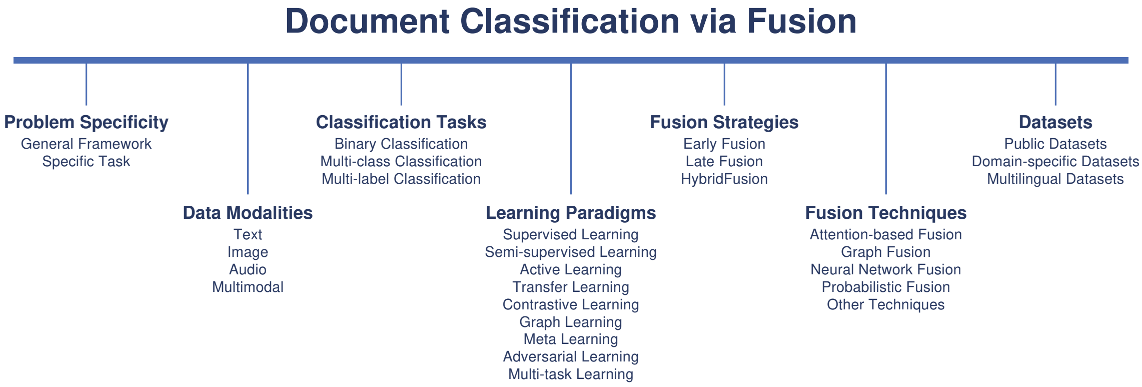

Figure 1: A taxonomy of document classification fusion strategies.

To clarify the scope of our analysis across different dimensions, we define the key aspects of the taxonomy as follows:

- Problem specificity: This dimension categorises studies based on whether the proposed method is designed for general-purpose document classification (foundational topics) or tailored to solve a specific, targeted problem (task-specific), such as sentiment analysis, topic identification, or fake news detection.
- Data modalities: This aspect distinguishes studies based on the types of data sources utilised. Common modalities involved in the reviewed works include text, image, audio, and video, often used in various combinations in multimodal approaches.
- Classification task: This dimension specifies the nature of the classification problem addressed, primarily distinguishing between binary classification (two output classes), multiclass classification (assigning one label from multiple mutually exclusive classes), and multilabel classification (assigning zero or more relevant labels from a predefined set).
- Learning paradigms: This dimension groups studies according to the underlying machine learning methodology used for training the classification models. Examples encountered include supervised learning, semi-supervised learning, active learning, transfer learning, contrastive learning, graph-based learning, meta-learning, adversarial learning, and multitask learning.
- Fusion strategies: This category concerns *when* information from different sources (views or modalities) is integrated relative to the main learning process. Key strategies include **early fusion** (combining input features or intermediate representations), **late fusion** (combining outputs or decisions from individual models), and **hybrid fusion** (employing a combination of early and late approaches).
- Fusion techniques: Distinct from fusion strategies, this aspect focuses on the particular *mechanisms* employed to combine information. The techniques identified by this review include attention mechanisms, graph-based methods (e.g. GNNs for fusion), dedicated neural network fusion layers, probabilistic

methods (e.g. Bayesian fusion), and classical machine learning algorithms adapted for fusion (e.g. kernel methods, specific ensemble techniques).

- Datasets: This category classifies studies based on the data corpora used for evaluation. This includes the use of established public benchmark datasets, newly curated custom datasets (often task- or language-specific), combinations or modifications of existing datasets, and multilingual datasets that are utilised frequently in multiview learning contexts.

6.1. Multimodal representation approaches

6.1.1. Problem specificity

The literature reveals a clear delineation based on problem specificity, with methods broadly categorised as either foundational topics (Table 11 and Figure 2) that provide general-purpose frameworks, or task-specific (Table 12 and Figure 3), meticulously designed for targeted applications. This distinction highlights how multimodal fusion techniques are developed with varying degrees of specialisation to address particular challenges and to leverage domain-specific nuances.

Table 11: Foundational and methodological works in multimodal learning.

Topics	**Works**
When to use multimodal learning	[80]
Feature selection as a multimodal optimisation problem	[81]
Handling missing modalities	[82]
Evaluating multimodal model performance	[83, 84]
Multimodal representation learning	[85, 86, 87, 88]
Multimodal data fusion	[89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124]

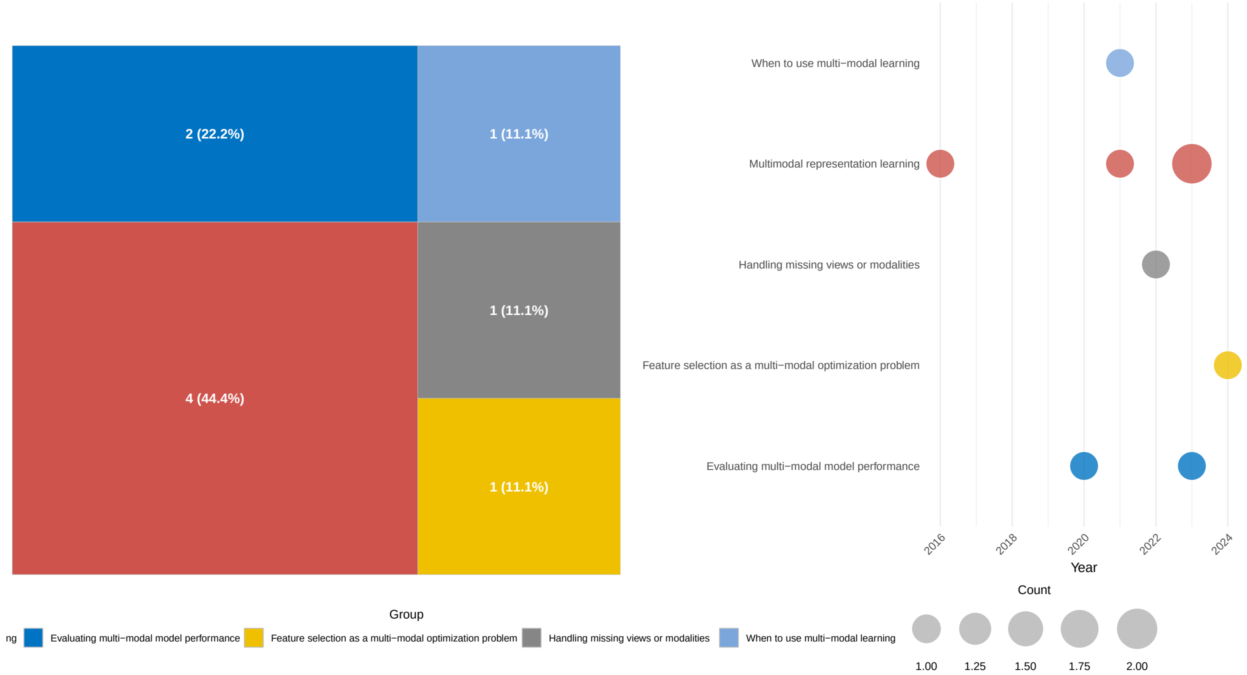

Figure 2: The left plot presents the number of publications broadly categorised as either foundational topics; the right plot depicts the time distribution of articles across sub-categories.

Figure 2 presents the distribution of foundational topics. Research in this area focuses on the foundational challenges inherent in multimodal learning, aiming for broader applicability. These studies explore robust multimodal representation learning techniques [86, 88], develop strategies for the handling of missing views or modalities [82], propose novel data fusion mechanisms [89, 93]), or investigate methods for feature selection, including multimodal optimisation approaches designed to identify diverse sets of optimal features [81].

The effectiveness of such general frameworks can be nuanced; for example, [80] observed that the utility of images in document classification can diminish when powerful pretrained language models are used, while [84] introduced diagnostics to determine whether performance gains truly stem from cross-modal interactions.

Table 12: Publications categorised as task-specific.

Topics	**Works**
Finance	[125]
Healthcare	[85, 126, 114]
Disaster response/managment	[96, 127, 105, 115]
Air traffic control	[97]
Cultural heritage	[82, 122]
Education	[102, 108]
E-commerce	[128, 129]
Industrial fault diagnosis	[106]
Social network	[104]
Social media analysis	[94, 98, 130, 110, 131, 132, 133]
Fake news detection	[99, 101]
Hate speech and offensive language detection	[130, 134]
Emotion recognition	[92, 100, 112, 119]
Author profiling	[131, 132, 117, 118, 120]
Document image classification	[135, 136, 125, 83, 103, 137, 107, 109, 113, 138]
Speaker role identification	[97]
Sentiment analysis	[89, 90, 95, 111, 139]
Filled pause detection	[140]
Safety report observations	[141]
Malware document classification	[142]
Research article classification	[91, 143]
Web document classification	[80, 121, 123]
Other multimodal tasks	[144, 145]

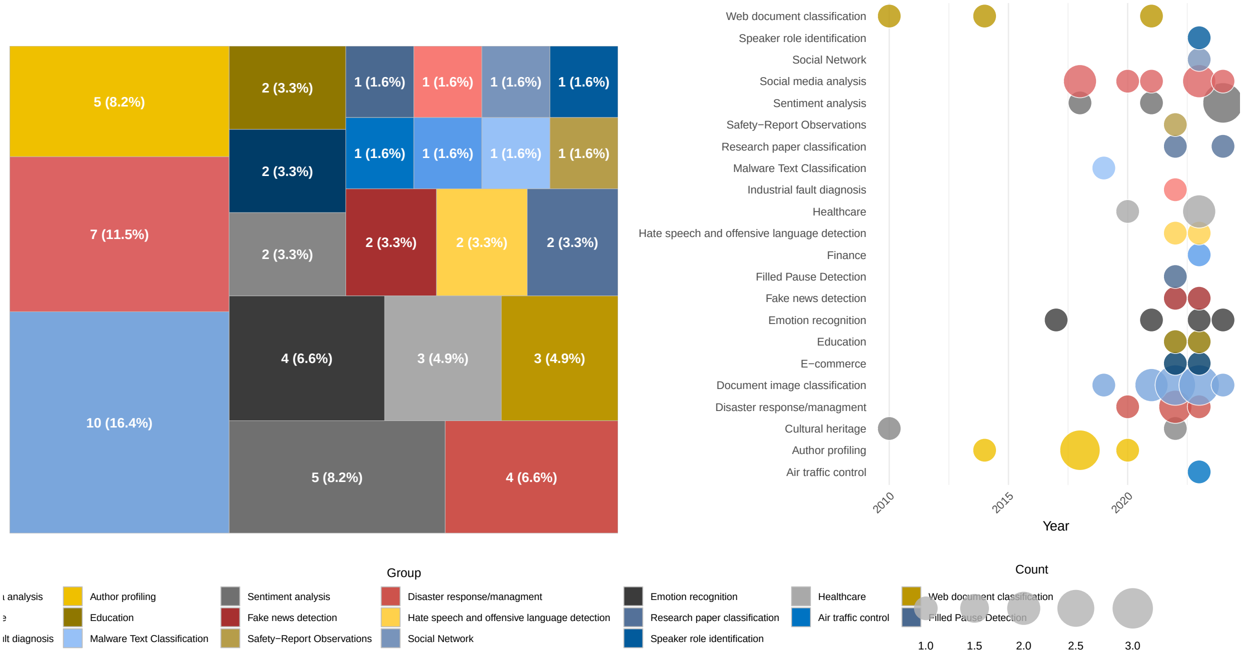

Figure 3: The left plot presents the number of publications broadly categorised as task-specific; the right plot depicts the time distribution of articles across sub-categories.

Figure 3 presents that task-specific approaches exhibit significant adaptation to unique domain requirements, often achieving enhanced performance by tailoring feature engineering and model architectures. For instance, in healthcare, researchers have developed systems like ImTeNet [114], which integrates image and text features for improved musculoskeletal abnormality detection.

In disaster management, methods such as the one proposed by [105] enhance situational awareness by fusing meteorological data with geolocated social media posts to identify flood victims, while [115] improves flood tweet classification by incorporating contextual hydrological information.

Fake news detection benefits from frameworks like FR-Detect [99], which incorporate publisher-related features alongside multimodal content analysis. Similarly, specialised approaches exist for multimodal sentiment analysis, in which models like [111] integrate text, image, and concept features from social media to better gauge public opinion, or for understanding financial documents, where robotic process automation is combined with multimodal models [125] for classification and information extraction.

6.1.2. Data modalities

Table 13 and Figure 4 present the diverse data modalities used in document classification. The predominant approach involves the fusion of *text and image* data, with 45 studies exploring that combination. The works span various applications, from general document image classification [138, 136] and sentiment analysis [89, 94] to specialised tasks like fake news detection [101] and multimodal disaster tweet classification [127, 96].

Table 13: Publications categorised based on data modalities.

Topics	Works
Text & image	[89, 90, 91, 93, 128, 94, 95, 96, 136, 85, 125, 98, 101, 102, 103, 137, 86, 127, 146, 129, 107, 141, 147, 143, 87, 109, 110, 80, 111, 112, 113, 114, 131, 84, 138, 142, 132, 133, 117, 118, 88, 120, 121, 123, 124]
Text & audio	[92, 97, 126, 140, 108, 116, 145]
Symbolic music & metadata	[122]
Text & metadata	[99, 104, 130, 105]
Text & image & audio	[144, 119]
Text & motion capture & audio	[100]
Text & time series	[106, 115, 139]
Text & image & metadata	[135, 83, 82, 134]

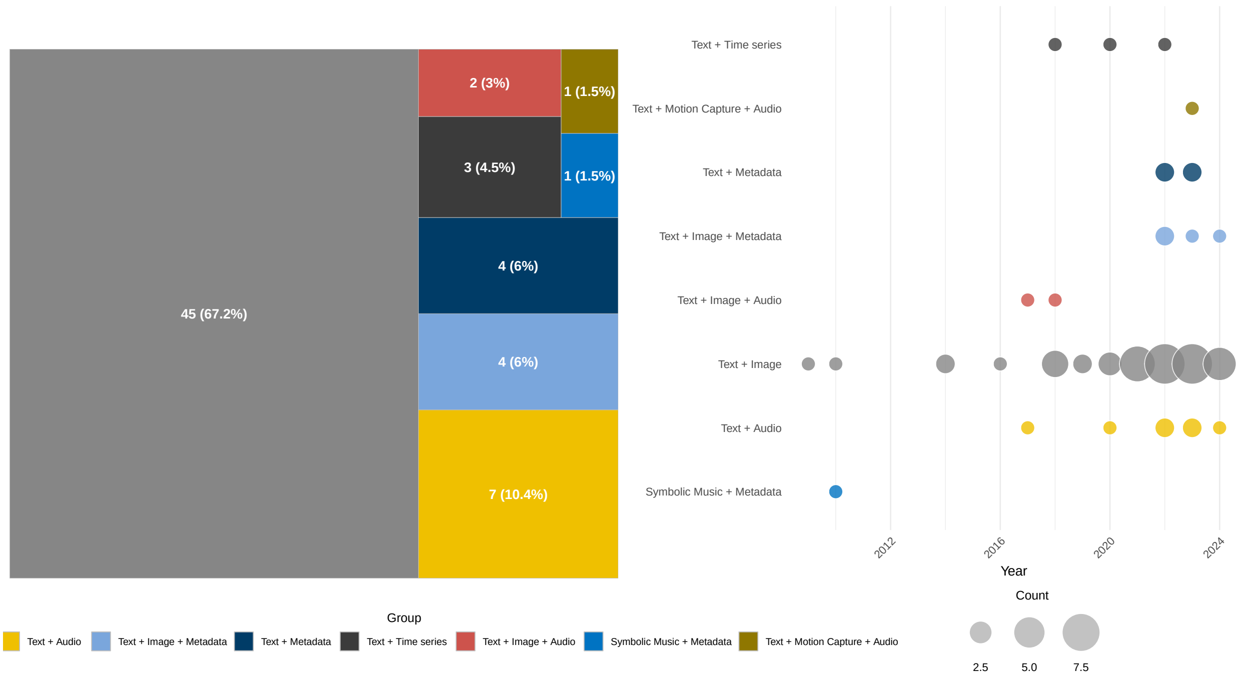

Figure 4: The left plot presents the number of publications broadly categorised within different data modalities; the right plot depicts the time distribution of articles across those modalities.

Methods that integrate *text and audio* are represented by eight of the studies that frequently target applications such as emotion recognition from speech and text [92], speaker role identification in specific contexts like air traffic control [97], the analysis of speech patterns for clinical applications like dementia prediction [126], and the analysis of speech patterns like filled pause detection [140]. The inclusion of *Metadata* alongside text is seen in six of the studies, which demonstrates a trend towards the leveraging of structured contextual information. This is evident in works that improve flood tweet classification by adding hydrological data [105], and approaches that improve the understanding of email reply behaviour by integrating social network dynamics into communication modelling. [104].

Fewer studies explore combinations of three modalities: *text, image and audio* are utilised by, for example [144] for addressee detection; *text, motion capture, and audio* modalities are used by, for example [100] for emotion recognition; *text, image and layout* modalities are used by, for example [83] for a multilingual document benchmark; and *text, image, and metadata* are used by [82] for classifying cultural heritage artifacts and [135] for classifying technical documents using network information. A unique approach that combines *text and time series data* was investigated by [106] to diagnose industrial faults using LSTMs, by [115] to improve flood tweet classification, and by [139] to improve polarity detection.

6.1.3. Classification tasks

The reviewed studies address a spectrum of classification tasks, primarily categorised by the nature of their output. These can be divided broadly into binary, multiclass, and multilabel classification, each reflecting different problem complexities and real-world applications (Table 14 and Figure 5).

Table 14: Publications categorised based on classification tasks.

Topics	**Works**
Binary classification	[97, 98, 99, 130, 101, 126, 104, 127, 140, 105, 134, 114, 115, 142, 144, 132, 133, 117, 118, 145]
Multiclass classification	[89, 135, 90, 91, 92, 93, 128, 94, 95, 136, 85, 125, 98, 100, 102, 103, 82, 86, 127, 146, 129, 106, 107, 141, 108, 147, 143, 87, 109, 111, 112, 113, 131, 84, 116, 138, 139, 119, 88, 120, 121, 122, 123, 124]
Multilabel classification	[83, 134, 110, 80]

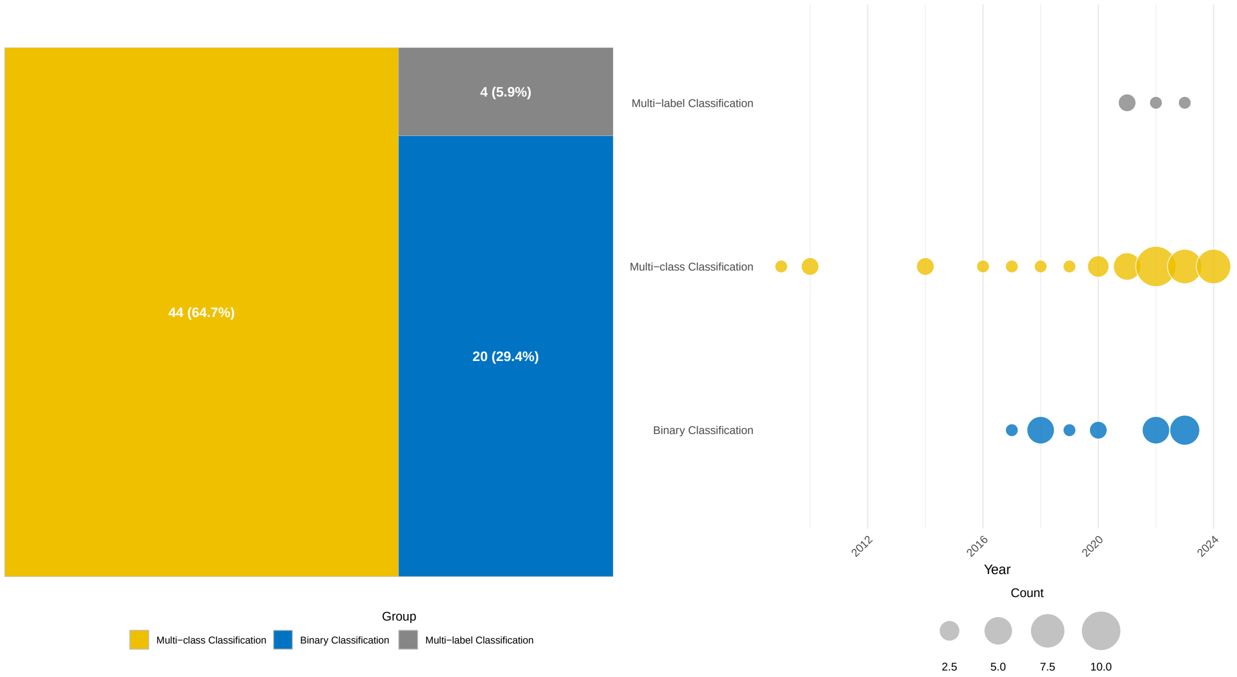

Figure 5: The left plot presents the number of publications broadly categorised within different classification tasks; the right plot depicts the time distribution of articles across the tasks.

Among the literature, *multiclass classification* is the most prevalent approach, with 45 studies. Those works aim to assign each document to one of several mutually exclusive categories. Examples range from the classification of emotions in multimodal data, in which [92] utilised a late fusion strategy for speech and text, to the categorisation of long documents, as seen in [93]'s hierarchical multimodal transformer, and the classification of pages of Brazilian legal documents [103]. The prominence of multiclass tasks suggests wide applicability in real-world scenarios in which documents must be sorted into distinct, predefined groups, often leveraging complex fusion strategies to differentiate nuanced categories. For instance, many articles in this category, such as [87] (visual word embedding for document classification) and [138] (multimodal document image classification), explore how combining visual and textual data enhances performance over unimodal approaches for these multicategory assignments.

Binary classification, addressed by 19 studies in this collection, focuses on the assignation of documents into one of two exclusive classes. This task is applied frequently to problems with clearly demarcated outcomes. For example, [97] focused on differentiating between air traffic controllers and pilots (a binary speaker role), [105] aimed to identify flood-affected areas or victims (presence/absence), and [101] tackled fake news detection (fake/real). These studies often highlight how the fusion of information fusion, such as the combination of meteorological data with social media posts [105] or image and text for fake news [101], leads to more precise and reliable binary decisions.

Multilabel classification, the least represented category with only four studies, tackles the most complex scenario, in which a document can be assigned zero, one, or multiple labels simultaneously from a predefined set. This reflects the inherent challenges in modelling potential label dependencies and the simultaneous relevance of multiple categories. The included studies, such as [83]'s work on multimodal multilingual document image benchmarks and [134]'s approach to multilabel misogyny detection, underscore the nascent but critical exploration of fusion techniques for these complex tasks. For instance, while also tackling a multilabel challenge, [134] employs fusion techniques (a multimodal ensemble) for an initial binary classification before using a text-only RoBERTa model to identify particular misogynistic labels.

6.1.4. Learning paradigms

The landscape of learning paradigms employed in information fusion-based document classification reveals distinct trends and evolving strategies. Table 15 and Figure 6 reveal that *supervised learning* dominates the field, with 49 distinct studies identified. This prevalence underscores its foundational role and efficacy, in which models learn from extensively labelled datasets to perform classification. These approaches range from sophisticated deep learning architectures that integrate multiple modalities, such as the use of meteorological data and social media text for enhanced flood detection [105] or the combination of text and visual features for on-device document classification [109], to novel transformer-based models designed for multimodal long document classification by capturing hierarchical structures and cross-modal associations [93].

Table 15: Publications categorised based on learning paradigms.

Topics	**Works**
Supervised learning	[89, 135, 90, 91, 93, 128, 94, 85, 97, 98, 99, 100, 130, 101, 102, 103, 126, 137, 104, 127, 146, 140, 105, 106, 134, 108, 143, 87, 80, 111, 112, 109, 114, 131, 115, 84, 138, 142, 144, 132, 117, 118, 145, 119, 120, 121, 122, 123, 124]
Semi-supervised learning	[129, 139, 133]
Active learning	[110]
Transfer learning	[89, 92, 85, 125, 83, 140, 107, 147, 113, 88]
Contrastive learning	[136, 86, 129, 141]
Multitask learning	[82]

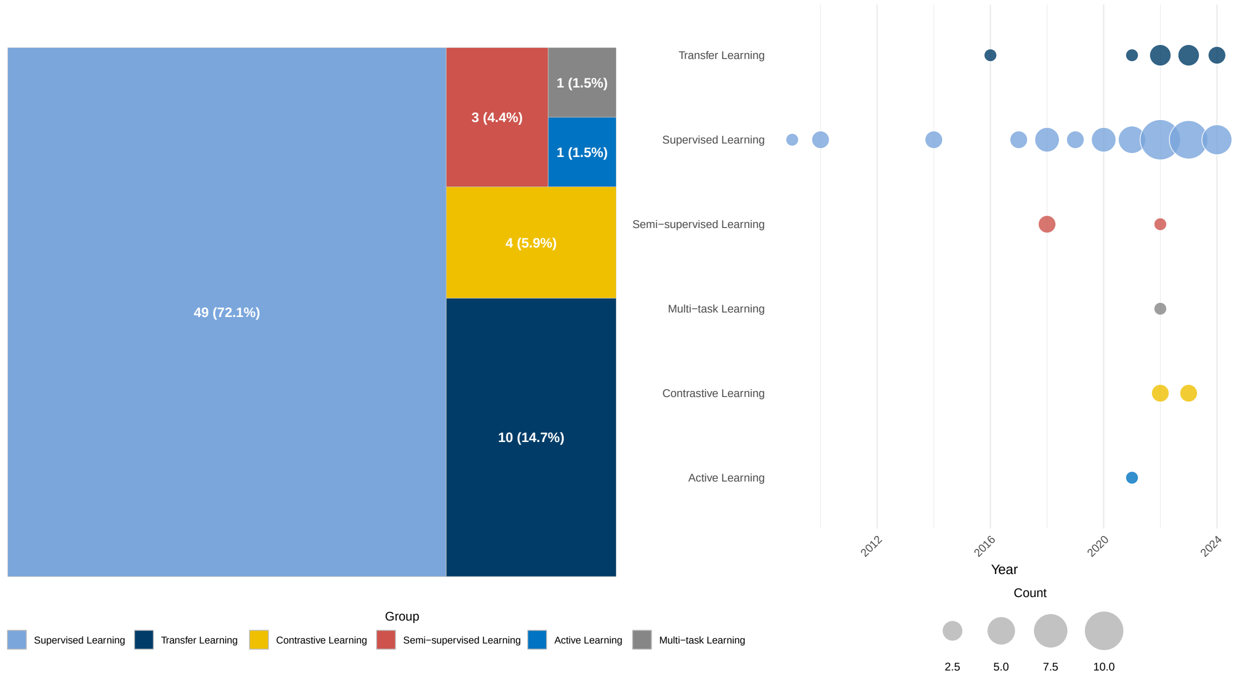

Figure 6: The left plot presents the number of publications categorised broadly within different learning paradigms; the right plot depicts the time distribution of articles across those paradigms.

Transfer Learning has emerged as a well-utilised paradigm, represented by 10 articles. This strategy is crucial in the leveraging of knowledge from pretrained models, applying it to new tasks or domains often characterised by limited data. For example, researchers have integrated capsule networks with deep ensemble learning and transfer learning for multimodal sentiment analysis, yielding substantial performance gains [89]. Other applications include the enhancement of document classification through side-tuning techniques that adapt base models without altering them [113], and the development of methods for heterogeneous domain adaptation by projecting text and image features into a shared subspace [147].

Growing interest is observed in *contrastive learning*, with four studies employing the approach to learn robust representations by contrasting similar and dissimilar data samples. This is exemplified by models like VLCDoC, which uses a cross-modal contrastive learning objective to align features from visual and textual modalities in document classification [136]. Similarly, cross-modal contrastive learning has been utilised to improve item categorisation by adapting BERT models with both text and image signals—notably without requiring images during the inference phase [129]—and to reduce decision bias caused by inconsistent unimodal information through contrastive alignment under weak supervision [86].

Semi-supervised learning, also represented by four articles, offers a valuable approach for scenarios with scarce labelled data by incorporating unlabelled examples into the training process. An illustrative example can be found in the development of a deep multimodal architecture for polarity detection in arguments, which leverages both labelled and unlabelled data effectively to build rich user representations [139].

Less frequently represented, yet important categories are *active learning* (one article) and *multitask learning* (one article). Active learning strategies, such as the ReALMS framework, aim to minimise the annotation burden by intelligently selecting the most informative samples for labelling, thereby improving model performance in multimodal document classification [110]. In the realm of multitask learning, one study demonstrates its application in the classification of multiple properties of historical silk fabric artifacts by leveraging image, text, and tabular data. This suggests potential benefits from the joint learning of related classification tasks [82].

6.1.5. Fusion strategies

The integration of information from diverse sources in document classification predominantly falls into three key strategies based on *when* fusion occurs relative to the main learning process: early, late, and hybrid fusion. Table 16 and Figure 7 presents the works and distributions along the fusion strategies.

Table 16: Publications categorised based on fusion strategies.

Topics	**Works**
Early fusion	[96, 135, 91, 92, 93, 128, 95, 136, 125, 97, 99, 100, 130, 101, 103, 126, 137, 104, 127, 105, 107, 141, 108, 147, 143, 87, 110, 80, 113, 114, 131, 115, 116, 138, 142, 139, 132, 133, 119, 120, 122]
Late fusion	[96, 90, 94, 82, 146, 140, 105, 106, 134, 112, 109, 111, 142, 117, 118, 121, 123, 124]
Hybrid fusion	[89, 85, 83, 102, 86, 105, 142, 144, 145, 88]

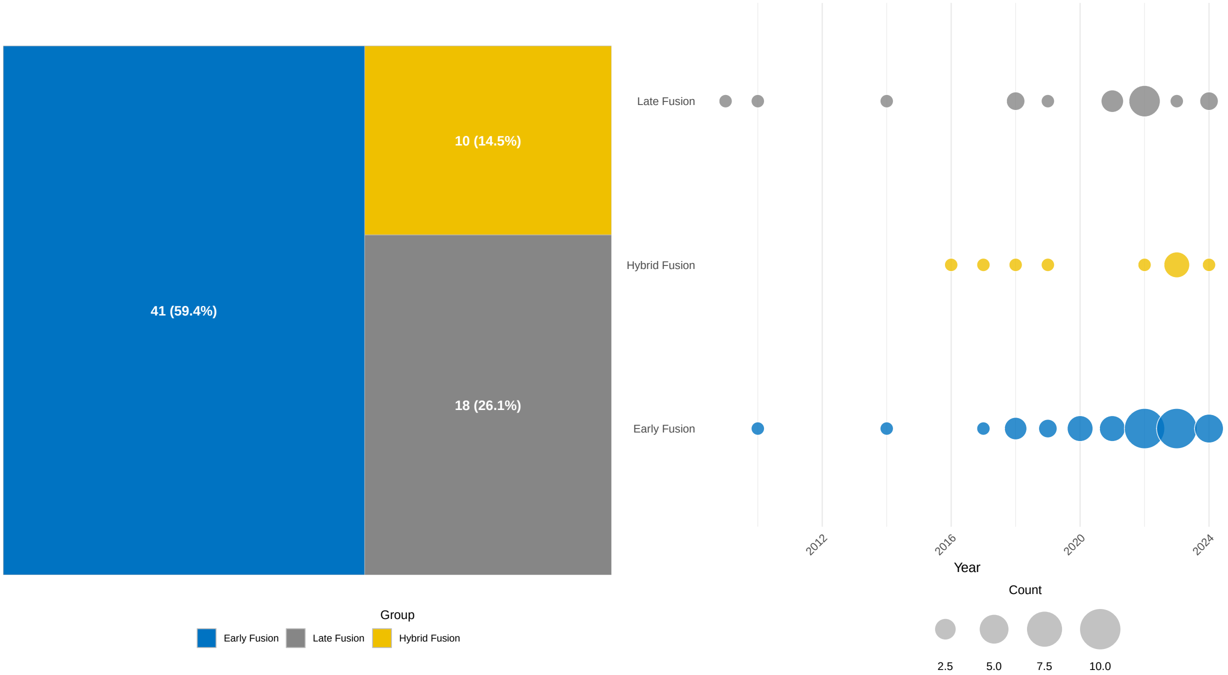

Figure 7: The left plot presents the number of publications broadly categorised within different fusion strategies; the right plot depicts the time distribution of articles across those strategies.

An analysis of the surveyed literature reveals *early fusion* to be the most common approach, utilised in approximately 58% of the studies. This strategy involves combining features at the input level or during initial representation stages. For instance, some methods transform text into visual representations, such as [87] which converted text into RGB images via Word2Vec for CNN processing. More advanced early fusion techniques include sophisticated feature integration, like the Cross-Modal Multiple Granularity Interactive Fusion Network in [91] for long document classification, or the use of intermodality, cross-attention, and intramodality self-attention modules as seen in [136].

Late fusion, which constitutes approximately 26% of the methods, combines outputs or decisions from models trained independently on different modalities. This approach offers flexibility, as exemplified by [111], in which predictions from text, image, and concept classifiers were combined using logistic regression as a metaclassifier for sentiment analysis. Similarly, [118] independently profiled text and image data from tweets for gender identification, subsequently merging the results. The work of [96] also demonstrates late fusion by combining outputs from text (LSTM with BERT) and image (CNN) classifiers for disaster response.

Hybrid fusion, employed in approximately 16% of the reviewed studies, represents a more complex integration, blending elements of both early and late approaches. For example, [89] explicitly used Yager fusion rules at both early and late stages for multimodal sentiment analysis. Another instance is [102], which combined models like CamemBERT and LayoutLMv2 through both early and late fusion techniques for noisy textbook exercise classification. The study [142], which is noted across all three fusion categories in this review's taxonomy, highlighted the potential for intricate strategies, in its case fusing QR code representations with text data for malware classification, which suggests a multifaceted integration that could be interpreted as hybrid.

6.1.6. Fusion techniques

Our categorisation identifies several key fusion techniques approaches: machine-learning-based fusion, probabilistic fusion, neural network fusion, and attention fusion (Table 17 and Figure 8).

Table 17: Publications categorised based on fusion techniques.

Topics	**Works**
Machine-learning-based fusion	[82, 105, 147, 111, 112, 131, 144, 132, 118, 145, 121, 123, 124]
Probabilistic fusion	[89, 106, 120, 122]
Neural network fusion	[135, 90, 91, 92, 93, 128, 94, 125, 83, 98, 99, 100, 130, 102, 103, 126, 104, 86, 127, 140, 107, 141, 108, 143, 87, 109, 110, 80, 113, 114, 115, 138, 142, 139, 133, 88]
Attention fusion	[136, 85, 97, 101, 137, 92, 116]

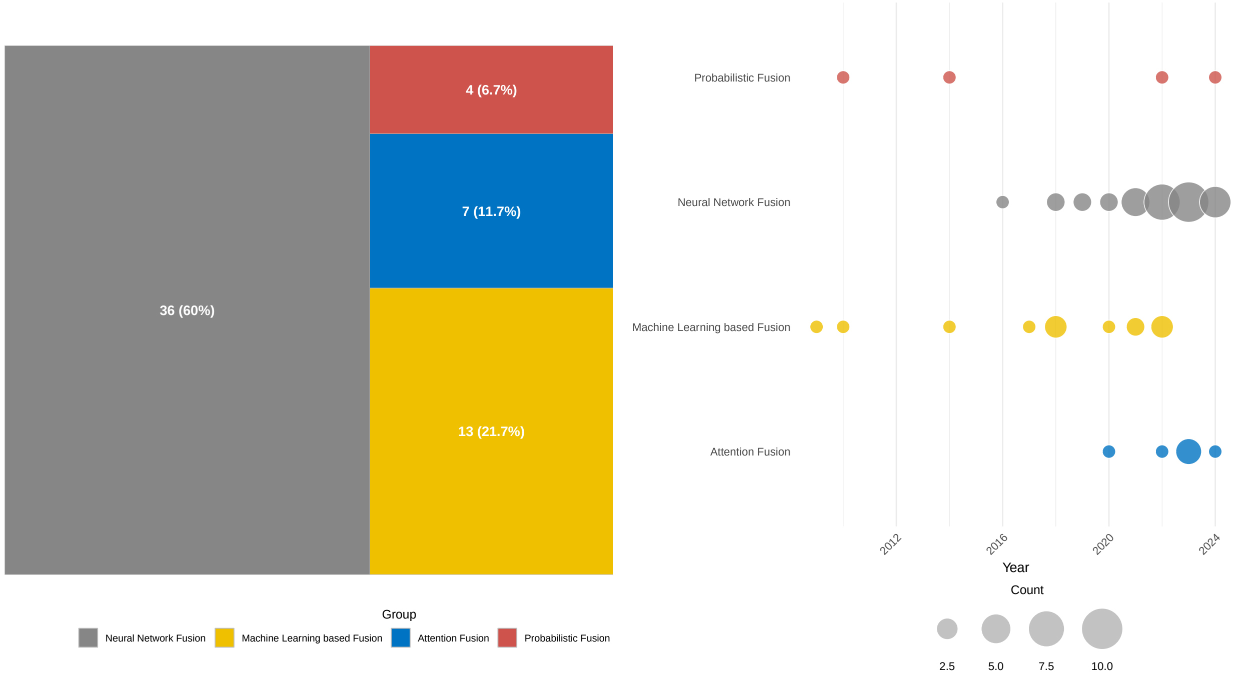

Figure 8: The left plot presents the number of publications categorised broadly within different fusion techniques; the right plot depicts the time distribution of articles across those techniques.

Machine-learning-based fusion adapts traditional algorithms to merge distinct information streams. For instance, [111] developed a framework for sentiment analysis that integrates text, image, and concept features using ensemble learning, employing logistic regression as a metaclassifier to combine predictions. Similarly, [105] demonstrated improved situational awareness in urban floods by fusing meteorological data with social media posts, while [147] introduced the Cross-Domain Structure Preserving Projection (CDSPP) algorithm to map heterogeneous features (e.g. text and images) into a shared subspace.

Probabilistic fusion methods provide a mathematically grounded framework for combining information, often addressing uncertainty. A multimodal sentiment analysis system that utilises Yager fusion rules to amalgamate outputs at both early and late fusion stages was proposed in [89] . In [106], a method for classifying multimodal heterogeneous data based on LSTM networks was presented, incorporating an adaptive weight fusion strategy for improved accuracy. Earlier work by [122] used a naïve Bayes classifier to fuse harmonic and instrumental features for the classification of music genres.

Dominating the landscape, *neural network fusion* leverages the potent representational capabilities of deep learning. There is a clear trend towards sophisticated multimodal architectures: [93], for example, introduced a hierarchical prompt and multimodal transformer (HPMT) that models interactions at both section and sentence levels within documents. In [95], it is showcased how LLMs like ChatGLM-6B can be enhanced for document classification by injecting image information as prompts, effectively fusing modalities with potentially fewer computational resources than some dedicated multimodal models.

A specialised and increasingly prominent subset of neural approaches is *attention fusion*, which dynamically weights the contributions of different information sources or features. In [136], VLCDoC, a cross-modal representation learning model that features intermodality cross-attention and intramodality self-attention modules, was proposed. In [85], PTUnifier, a model for medical vision-and-language pretraining that uses visual and textual soft prompts to unify different pretraining paradigms, was developed. In [97], a design of MMSRINet for speaker role identification in air traffic communication by fusing speech and text features using attention mechanisms was proposed. In [116], a method for synchronised word representation that incorporates phonetic information alongside text through an attention mechanism is proposed.

6.1.7. Datasets

Our analysis of the literature, based on the provided taxonomy's classifications, identifies three principal categories of dataset (Table 18 and Figure 9): public benchmark datasets (comprising approximately 29% of the cited studies), domain-specific datasets (the largest group at approximately 64%), and multilingual datasets (a focused 7%).

Table 18: Publications categorised based on datasets.

Topics	**Works**
Public benchmark datasets	[89, 90, 91, 93, 128, 94, 136, 100, 130, 101, 86, 129, 107, 147, 87, 113, 138, 132, 117, 118, 88, 124]
Domain-specific datasets	[135, 92, 95, 96, 85, 125, 83, 97, 98, 99, 102, 103, 126, 137, 82, 104, 127, 146, 140, 129, 105, 106, 134, 141, 108, 143, 109, 112, 110, 111, 114, 131, 115, 116, 142, 144, 139, 133, 145, 119, 120, 121, 122, 123]
Multilingual datasets	[125, 83, 101, 80, 132, 117]

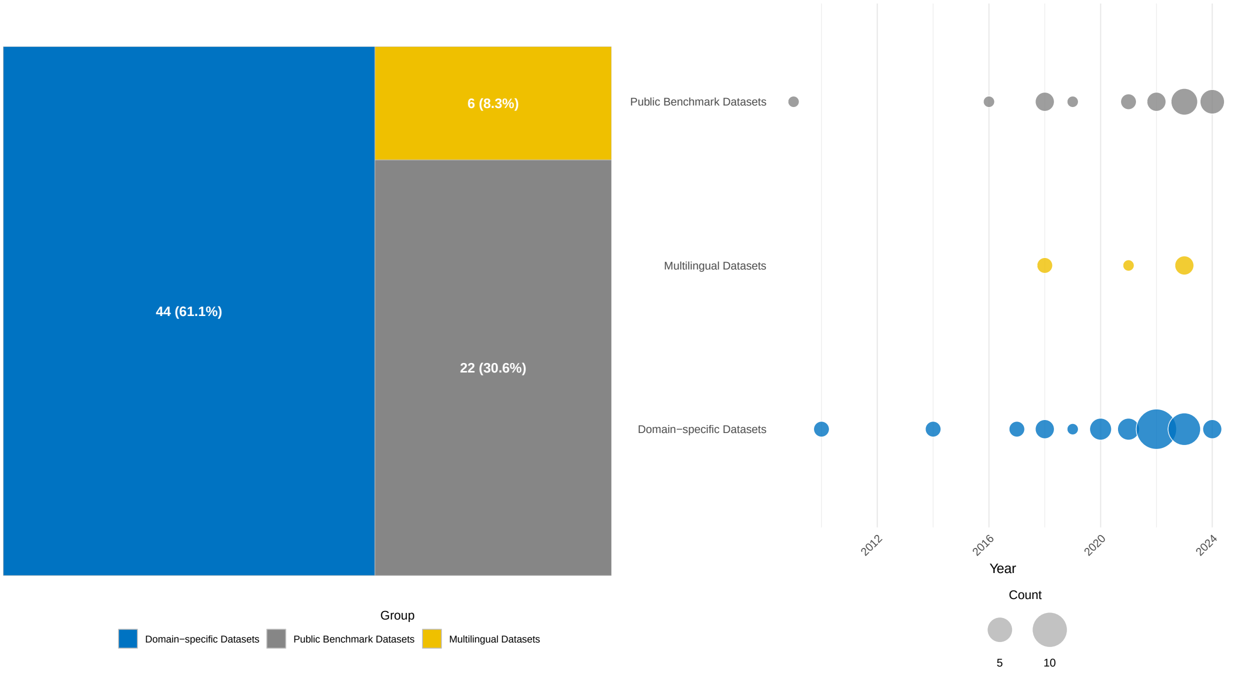

Figure 9: The left plot presents the number of publications broadly categorised within different types of datasets; the right plot depicts the time distribution of articles across those types.

Public benchmark datasets are foundational for standardised evaluation, enabling researchers to compare the performance of novel methods directly. Prominent examples include RVL-CDIP for document image classification, utilised in studies like [136] that investigated vision–language contrastive pretraining, and social media collections such as Twitter-15 and Twitter-17, employed by [94] for multimodal classification with compact bilinear pooling. These benchmarks have been instrumental in the validation of innovative architectures, such as hierarchical prompt and multimodal transformers for long document classification [93], and visual word embeddings for document classification [87].

Domain-specific datasets represent the most substantial portion of the surveyed research. This underscores the increasing application of fusion techniques to address nuanced challenges in specialised fields. These custom-curated datasets cater to particular requirements in areas like healthcare (e.g. [85] unifying medical vision-and-language pretraining), disaster management (e.g. [105] improving flood awareness with multimodal fusion, legal document analysis [103], and cultural heritage ([82] classifying silk fabric artifacts). The creation of such datasets often involves tackling inherent difficulties such as noisy or unbalanced data and limited sample sizes. This was highlighted by [102] in the context of classifying French textbook exercises.

Multilingual datasets constitute a smaller yet pivotal category, essential in the development of globally relevant and robust classification models. Research in this area frequently explores cross-lingual transfer learning and the unique challenges of fusing information from different languages or modalities across linguistic boundaries. For instance, [83] introduced the WikiDoc and MultiEURLEX datasets as multilingual benchmarks for document image classification, while [80] investigated the (in)effectiveness of images for document classification across multiple languages. Studies like [125] also demonstrate the application of multimodal approaches to multilingual financial documents.

6.2. Multiview representation approaches

6.2.1. Problem specificity

Our analysis of information fusion-based document classification methods reveals a notable division between general-purpose frameworks (58% of the studies, Table 19, Figure 10) and task-specific approaches (42%, Table 20, Figure 11). This distribution indicates a field in which methodological innovations and practical applications evolve in parallel.

Table 19: Foundational and methodological works in multiview learning.

Topics	**Works**
When to use multiview learning	[148, 149]
Handling missing views	[150, 151, 149]
Multiview representation learning	[152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175]
Multiview data fusion	[176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203]

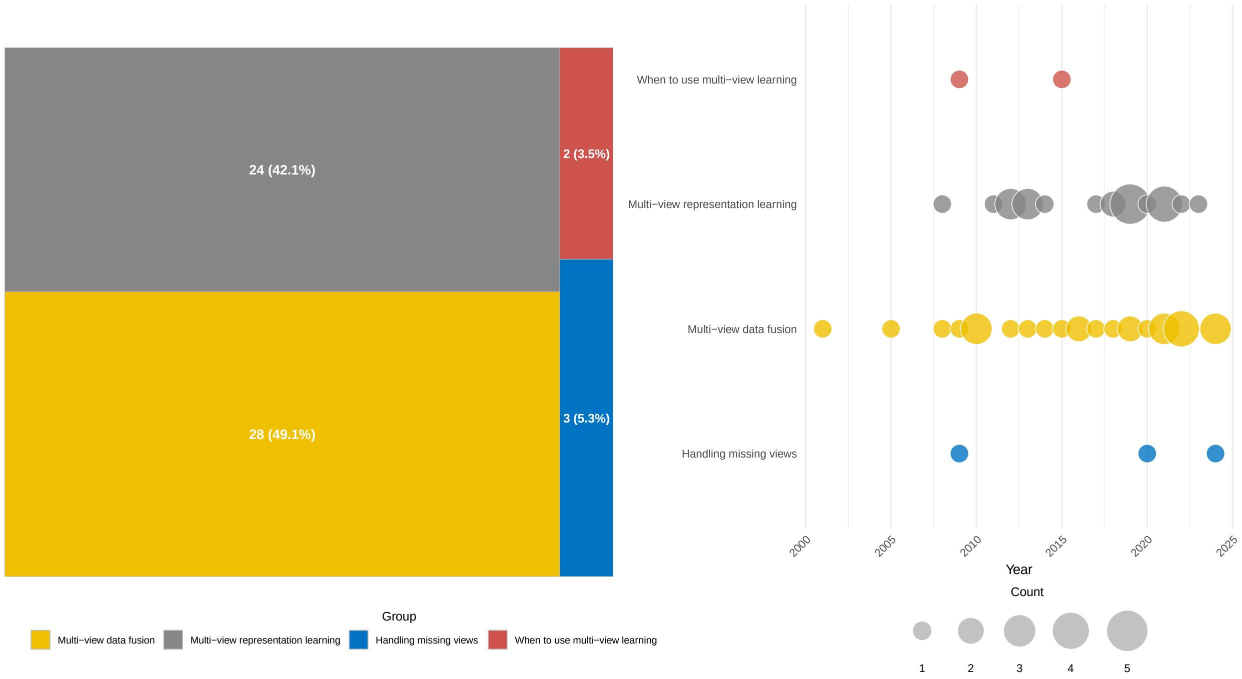

Figure 10: The left plot presents the number of publications categorised broadly as foundational topics; the right plot depicts the time distribution of the articles across sub-categories.

General frameworks focus predominantly on multiview representation learning (28 studies) and multiview data fusion (27 studies), with smaller but important contributions addressing missing views, feature selection, and theoretical foundations. Recent innovations in this category include the deep incomplete multiview learning network for handling missing data by [150] and adaptive knowledge incorporation techniques that enhance classification through multiview feature fusion by [176].

Table 20: Publications categorised as task-specific.

Topics	**Works**
Finance	[180]
Social media analysis	[204, 205, 206]
Hate speech and offensive language detection	[204, 207]
Fake news detection	[179]
Emotion recognition	[158]
Sentiment analysis	[158, 163]
Author profiling	[131]
Technology forecasting	[184]
Multilingual text categorisation	[156, 151, 159, 161, 167, 88, 193, 171, 172, 197, 198, 149]
Adversarial document classification	[207]
Lobbying disclosure analysis	[208]
Software document classification	[194]
Short document classification	[183, 195]
Web document classification	[187, 188, 199, 209, 210, 211, 201]
News article classification	[203]
Biomedical document, article analysis	[181]
Healthcare	[182, 190, 192]
Personality prediction	[153]
Extreme multilabel document classification	[162]
Other multiview tasks	[212, 165]

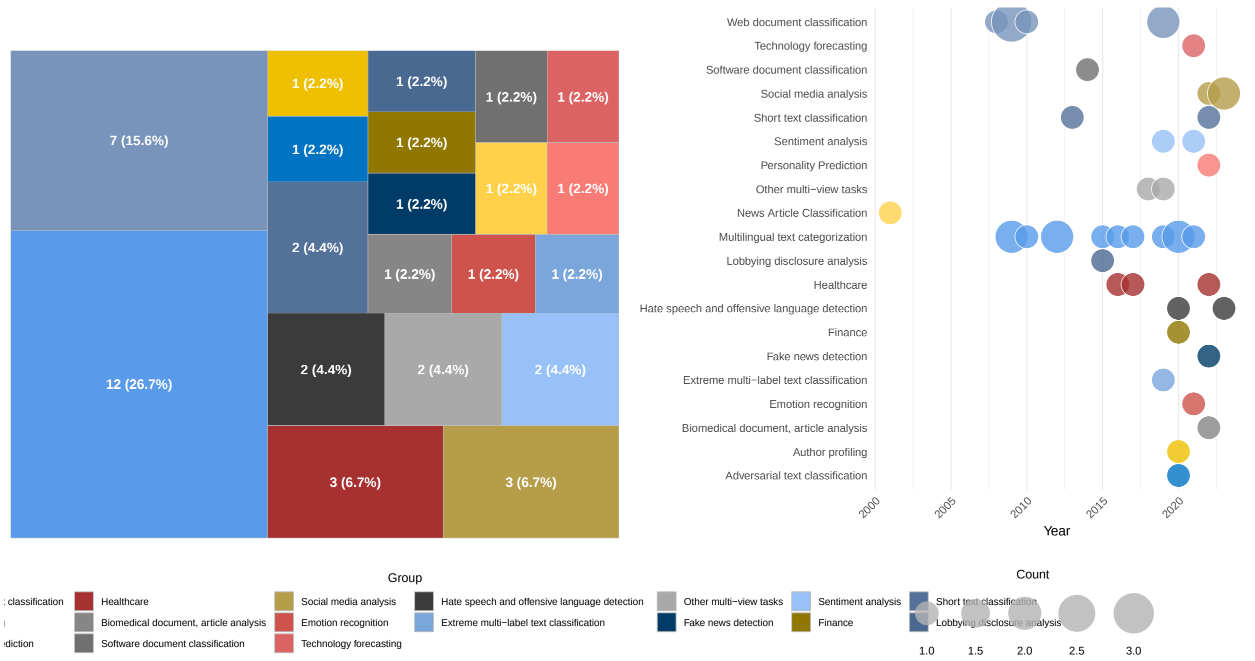

Figure 11: The left plot presents the number of publications categorised broadly as task-specific; the right plot depicts the time distribution of the articles across sub-categories.

Task-specific approaches demonstrate remarkable diversity across 19 different application domains. Multilingual text categorisation (11 studies) and web document classification (nine studies) represent the most extensively explored areas—likely due to their inherently multiview nature. The multilingual domain has fostered innovations in cross-lingual information fusion, as seen in the work of [151] on handling missing views in multilingual contexts. Recent years have witnessed an expansion into socially relevant domains including social media analysis [204, 205], hate speech detection [207], and fake news detection ([179], who fused global and local semantic views through BERT and CNN). Healthcare applications have also gained traction, with [182] addressing the challenges of identifying adverse drug reactions through multiview active learning.

6.2.2. *Classification tasks*

Our taxonomy of information fusion-based document classification methods reveals an uneven distribution of research across different classification paradigms (Table 21 and Figure 12).

Table 21: Publications categorised based on classification tasks.

Topics	Works
One-class classification	[185, 209]
Binary classification	[204, 205, 179, 181, 206, 182, 153, 131, 207, 187, 212, 188, 189, 190, 213, 148, 194, 157, 195, 169, 170, 171, 196, 173, 197, 198, 199, 149, 201, 175, 202]
Multiclass classification	[177, 178, 150, 152, 180, 183, 184, 154, 155, 156, 158, 151, 159, 214, 160, 215, 163, 164, 165, 216, 167, 217, 88, 191, 192, 193, 218, 168, 172, 174, 200, 210, 211, 203]
Multilabel classification	[176, 186, 158, 162, 166, 208]

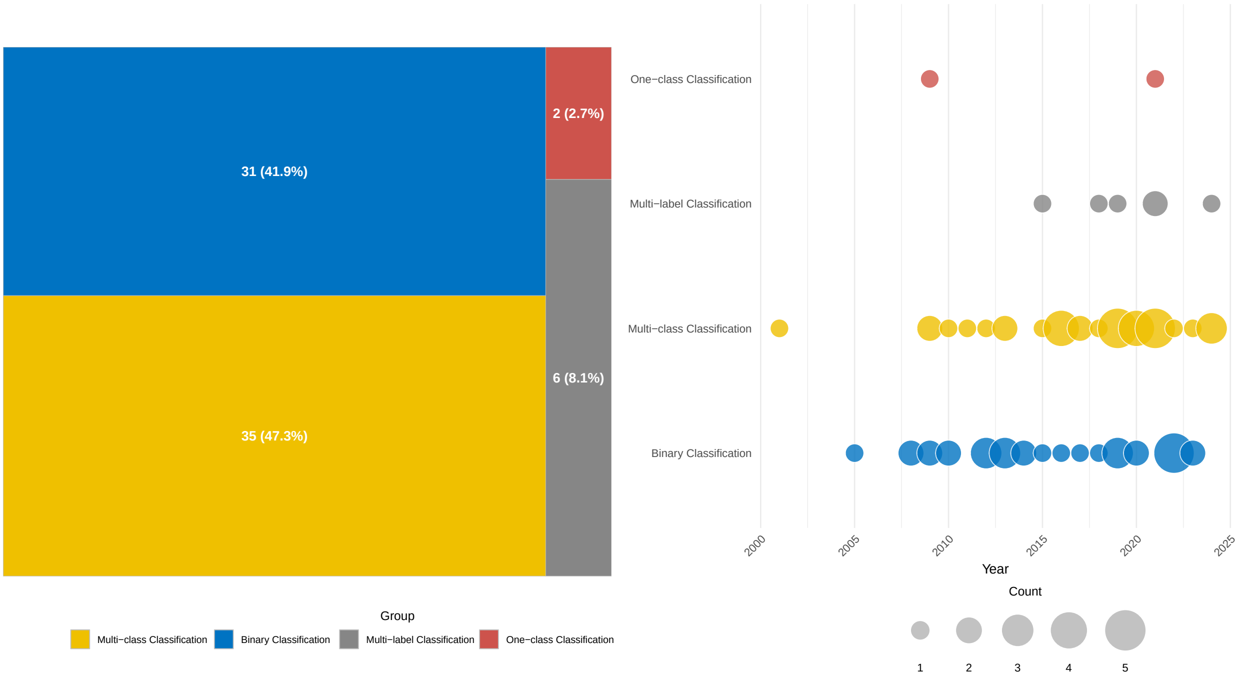

Figure 12: The left plot presents the number of publications categorised broadly within different classification tasks; the right plot depicts the time distribution of articles across those tasks.

Multiview multiclass classification dominates with 36 articles spanning diverse applications from document categorisation [203, 210, 211] to technology forecasting [184] and multilingual text analysis [88]. This category exhibits considerable methodological diversity, from early information fusion techniques to recent transformer-based architectures with attention mechanisms [177, 178].

Similarly well represented is *multiview binary classification* with 30 articles, which focuses primarily on sentiment analysis [207], fake news detection [179], and author profiling [131]. Both categories demonstrate a clear methodological evolution from traditional machine learning approaches towards sophisticated deep learning architectures.

In contrast, *multiview multilabel classification* remains relatively underexplored with only six articles, despite addressing important challenges like hierarchical document classification [176] and extreme multilabel scenarios [162].

The smallest category, *multiview one-class classification* (two articles), introduces specialised approaches like multiview deep support vector data description [185] for scenarios in which only positive examples are available.

6.2.3. Learning paradigms

The taxonomy of learning paradigms in information-fusion-based document classification reveals a rich landscape of methodological approaches, each leveraging multiple views to enhance classification performance (Table 21, Figure 13).

Table 22: Publications categorised based on learning paradigms.

Topics	**Works**
Supervised learning	[176, 204, 179, 180, 181, 183, 184, 155, 186, 156, 158, 131, 187, 212, 161, 162, 188, 165, 189, 166, 190, 216, 167, 191, 192, 208, 194, 218, 170, 172, 174, 197, 200, 149, 203]
Semi-supervised learning	[185, 177, 150, 154, 159, 217, 213, 148, 193, 195, 171, 196, 198, 149, 209, 210, 201, 175, 202]
Active learning	[177, 206, 182, 199, 210, 211, 201]
Transfer learning	[176, 161, 215, 88, 157, 169, 173]
Contrastive learning	[152]
Graph learning	[178]
Meta learning	[205]
Few-shot learning	[205]
Adversarial learning	[151, 207]

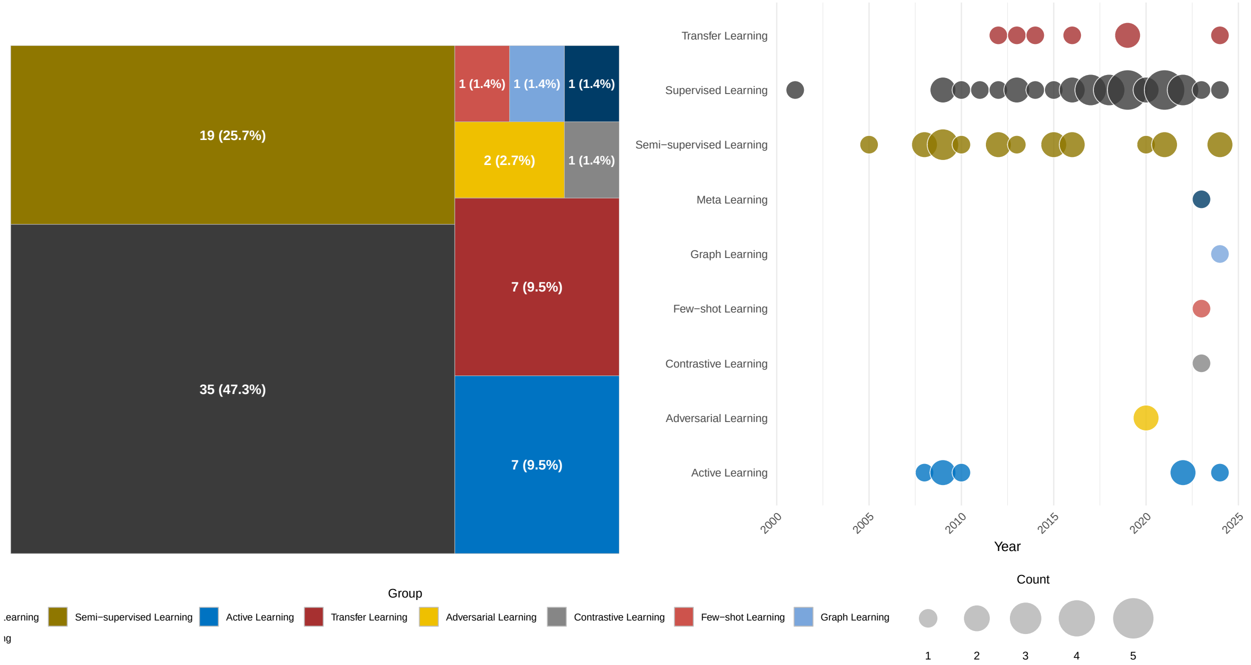

Figure 13: The left plot presents the number of publications categorised broadly within different learning paradigms; the right plot depicts the time distribution of articles across those paradigms.

The chronological distribution of research demonstrates a progressive shift from traditional supervised approaches towards more sophisticated paradigms. Early works (2001–2015) focused primarily on basic fusion techniques within supervised and semi-supervised frameworks, while recent research (2020–2024) explores advanced paradigms like few-shot and contrastive learning. This evolution responds to persistent challenges in document classification, including data scarcity, domain adaptation needs, and the complexity of capturing semantic relationships across multiple text representations.

The application domains span diverse areas including sentiment analysis, fake news detection [179], biomedical document classification [181], and technology forecasting [184]. Notably, models that adapt dynamically to particular domains through techniques like metalearning [205] and transfer learning [161, 215] demonstrate particular promise for specialised document classification tasks.

Looking forward, the emerging interest in hybrid paradigms—such as the combination of active and semi-supervised learning [177]—suggests that future research will increasingly blur traditional paradigmatic boundaries to create more adaptive, data-efficient classification frameworks for complex documents.

6.2.4. Fusion strategies

This section examines various approaches to information fusion in document classification, categorised by *when* integration occurs in the learning process. The taxonomy reveals distinct patterns across 69 studies spanning over two decades of research (Table 23, Figure 14).

Table 23: Publications categorised based on fusion strategies.

Topics	**Works**
Early fusion	[177, 178, 150, 152, 179, 183, 153, 185, 154, 155, 186, 156, 158, 131, 207, 214, 160, 161, 164, 165, 190, 216, 167, 192, 218, 169, 196, 174, 203]
Late fusion	[204, 205, 181, 184, 187, 212, 163, 188, 189, 166, 217, 191, 148, 193, 208, 195, 172, 199, 200, 211, 201, 175, 202]
Hybrid fusion	[176, 180, 206, 182, 151, 159, 162, 215, 88, 213, 194, 170, 171, 197, 198, 209, 210]

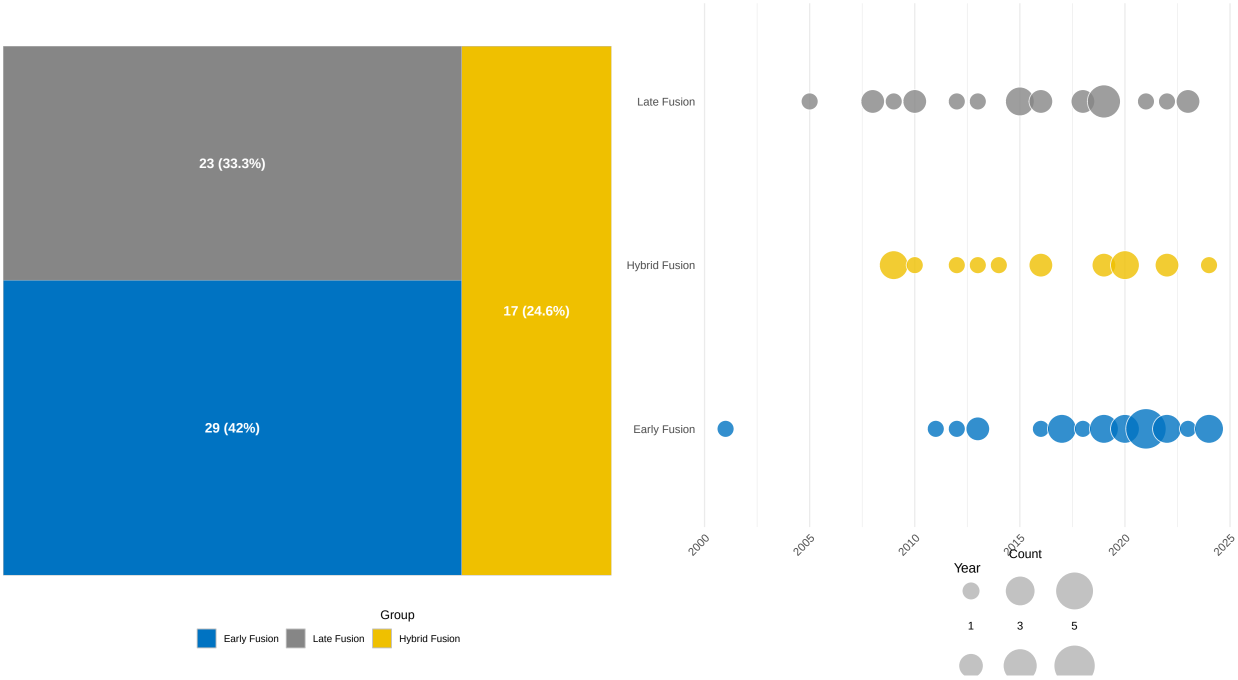

Figure 14: The left plot presents the number of publications categorised broadly within different fusion strategies; the right plot depicts the time distribution of articles across those strategies.

Multiview early fusion represents the most prevalent approach, in which integration occurs at the feature or representation level before the main classification decision. Recent trends in this category demonstrate a progression from traditional techniques like canonical correlation analysis [156, 169] towards sophisticated deep learning architectures. Many recent works leverage transformer-based models like BERT combined with complementary features [179, 178], while attention mechanisms have become increasingly popular for weighting different views [155, 158]. Notable innovations include multimodal embeddings that capture different aspects of text data simultaneously [161, 207] and frameworks that are designed specifically to handle incomplete views [150].

Multiview late fusion approaches integrate at the decision level, combining outputs from separate classifiers. This category features significant work on ensemble methods and metaclassification strategies [188, 187]. Many studies focus on cross-lingual and multilingual document classification [205, 172], in which separate models for each language are fused at decision time. Recent innovations include strategies that incorporate uncertainty measures when combining classifier decisions [204].

Multiview hybrid fusion represents the smallest but perhaps most sophisticated category, employing fusion at multiple levels of the classification pipeline. These approaches often address complex scenarios such as limited labelled data through active learning [206, 182] or semi-supervised techniques [159]. Many hybrid methods dynamically adjust fusion strategies based on data characteristics—particularly when handling missing views through generative models [151]—or on balancing the contribution of different views through tunable hyperparameters in a transfer learning context [215]—particularly in challenging contexts like extreme multilabel classification [162]. Recent innovations also include adaptive fusion mechanisms [176].

6.2.5. Fusion techniques

Table 24 and Figure 15 present the taxonomy of fusion techniques in document classification, which reveals distinct methodological patterns for integrating multiview information.

Table 24: Publications categorised based on fusion techniques.

Topics	**Works**
Machine-learning-based fusion	[204, 180, 181, 185, 156, 131, 187, 160, 212, 161, 215, 188, 164, 189, 166, 167, 191, 213, 195, 169, 170, 196, 174, 199, 211, 201]
Probabilistic fusion	[176, 163, 217, 148, 193, 168, 197, 198, 200, 210, 175]
Neural network fusion	[177, 150, 205, 179, 182, 183, 184, 154, 186, 151, 207, 214, 162, 190, 216, 88, 192, 203]
Attention fusion	[153, 155, 158]
Graph fusion	[176, 177, 178, 152, 196]

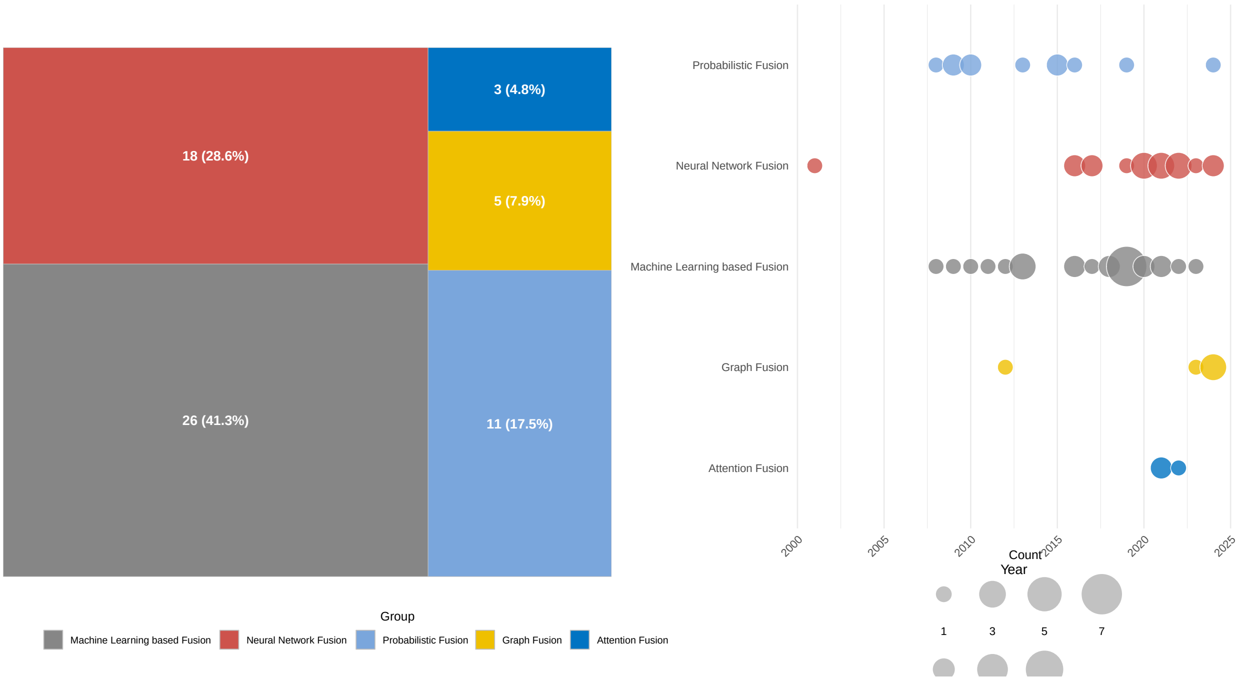

Figure 15: The left plot presents the number of publications categorised broadly within different fusion techniques; the right plot depicts the time distribution of articles across those techniques.

Machine-learning-based fusion represents the largest and most established category, encompassing a wide array of approaches from classical techniques like canonical correlation analysis (e.g. [156, 169]) and stacked generalisation [187] to more modern strategies that involve active learning [199, 211] and transfer learning [215].

A parallel stream, *probabilistic-based fusion*, provides strong theoretical foundations by employing Bayesian frameworks [163] and Dempster–Shafer theory for combining evidence [200]. This type of fusion excels at handling uncertainty.

The evolution towards deep learning is prominent in *neural-network-based fusion*, which exhibits a clear trajectory from early neural architectures [203] to sophisticated models that fuse features from distinct neural architectures [183, 216], leverage pretrained transformers [179], or learn disentangled shared and specific representations [154].

Though less prevalent, *attention-based fusion* offers unique advantages by introducing dynamic weighting mechanisms that focus selectively on salient features from different views, enhancing both performance and model explainability [155, 158].

Finally, *graph-based fusion* is emerging as a powerful and recent direction, effectively modelling complex structural relationships between views using GNNs and heterogeneous graph attention mechanisms, often in semi-supervised contexts [176, 177, 178].

6.2.6. Datasets

Our analysis of information-fusion-based document classification research reveals three distinct dataset categories (Table 25, Figure 16).

Table 25: Publications categorised based on datasets.

Topics	Works
Benchmark datasets	[176, 177, 178, 150, 152, 179, 183, 185, 154, 155, 186, 151, 214, 160, 161, 162, 215, 163, 164, 165, 189, 166, 216, 217, 88, 191, 192, 213, 148, 193, 157, 194, 218, 195, 168, 169, 170, 173, 174, 197, 198, 149, 209, 210, 211, 175, 202, 203]
Domain-specific datasets	[204, 180, 181, 182, 153, 184, 131, 207, 187, 212, 188, 190, 208, 199, 201]
Multilingual datasets	[205, 156, 131, 159, 161, 167, 171, 172, 197, 198, 149]

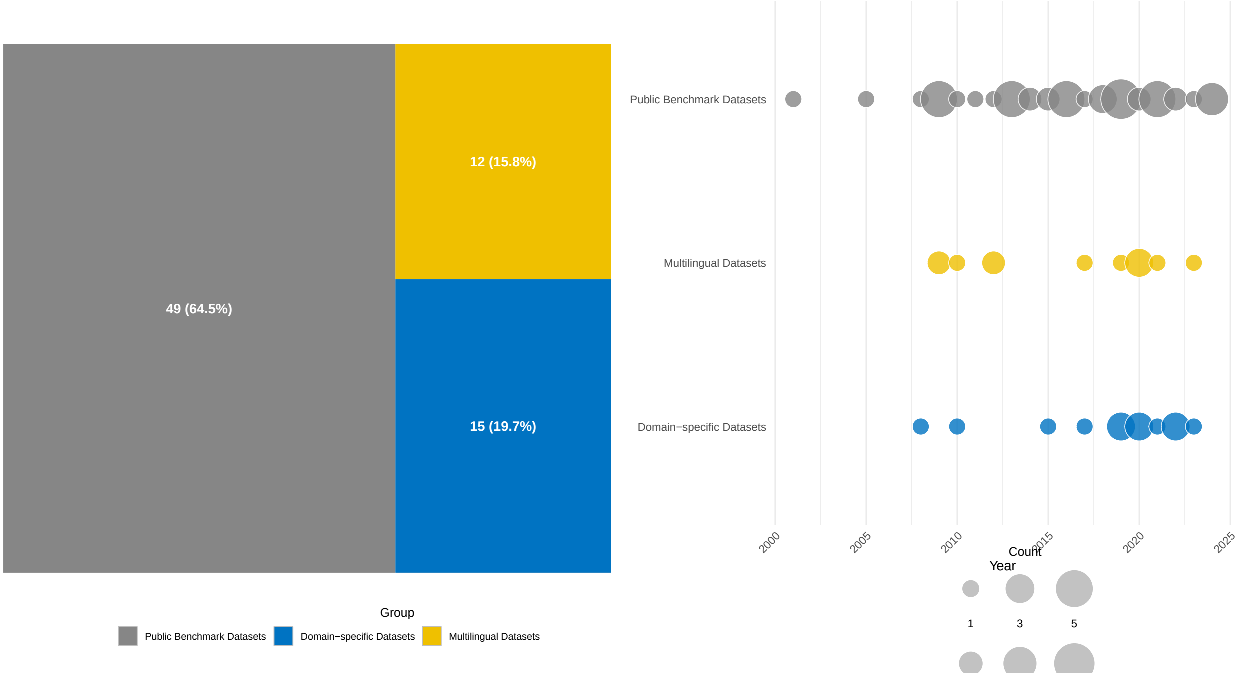

Figure 16: The left plot presents the number of publications categorised broadly within different types of datasets; the right plot depicts the time distribution of articles across those types.

First, *benchmark datasets*—predominantly the Reuters collections and 20 Newsgroups—serve as standardised evaluation platforms that enable consistent comparison across fusion methodologies. Second, *domain-specific datasets* address specialised applications in healthcare [190]), finance [180], and technology forecasting [184], often driving innovations in fusion techniques tailored to particular domains. Third, *multilingual datasets*, particularly the Reuters RCV1/RCV2 extensions [149, 197], facilitate cross-lingual fusion research in which each language represents a distinct view.

Dataset characteristics influence fusion method design heavily: benchmark datasets facilitate attention mechanism development [155], domain-specific collections inspire specialised fusion techniques [182], and multilingual corpora advance cross-lingual approaches [156].

References

[1] H. Zaheer, M. Bashir, Detecting fake news for covid-19 using deep learning: a review, Multimedia Tools and Applications 83 (2024) 74469–74502. doi:10.1007/s11042-024-18564-7.

[2] R. Anggrainingsih, G. Hassan, A. Datta, Transformer-based models for combating rumours on microblogging platforms: a review, Artificial Intelligence Review 57 (2024). doi:10.1007/s10462-024-10837-9.

[3] S. Singh, W. Singh, Ai-based personality prediction for human well-being from text data: a systematic review, Multimedia Tools and Applications 83 (2024) 46325–46368. doi:10.1007/s11042-023-17282-w.

[4] S. Biradar, S. Saumya, A. Chauhan, Combating the infodemic: Covid-19 induced fake news recognition in social media networks, Complex and Intelligent Systems 9 (2023) 2879–2891. doi:10.1007/s40747-022-00672-2.

[5] D. Rogers, A. Preece, M. Innes, I. Spasić, Real-time text classification of user-generated content on social media: Systematic review, IEEE Transactions on Computational Social Systems 9 (2022) 1154–1166. doi:10.1109/TCSS.2021.3120138.

[6] Y. HaCohen-Kerner, Survey on profiling age and gender of text authors, Expert Systems with Applications 199 (2022). doi:10.1016/j.eswa.2022.117140.

[7] N. Mullah, W. Zainon, Advances in machine learning algorithms for hate speech detection in social media: A review, IEEE Access 9 (2021) 88364–88376. doi:10.1109/ACCESS.2021.3089515.

[8] W. Yin, A. Zubiaga, Towards generalisable hate speech detection: a review on obstacles and solutions, PeerJ Computer Science 7 (2021) 1–38. doi:10.7717/PEERJ-CS.598.

[9] M. Al-Garadi, M. Hussain, N. Khan, G. Murtaza, H. Nweke, I. Ali, G. Mujtaba, H. Chiroma, H. Khattak, A. Gani, Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges, IEEE Access 7 (2019) 70701–70718. doi:10.1109/ACCESS.2019.2918354.

[10] J. Duarte, L. Berton, A review of semi-supervised learning for text classification, Artificial Intelligence Review 56 (2023) 9401–9469. doi:10.1007/s10462-023-10393-8.

[11] F. Elsafoury, S. Katsigiannis, Z. Pervez, N. Ramzan, When the timeline meets the pipeline: A survey on automated cyberbullying detection, IEEE Access 9 (2021) 103541–103563. doi:10.1109/ACCESS.2021.3098979.

[12] I. Botunac, M. B. Bakarić, M. Matetić, Comparing fine-tuning and prompt engineering for multi-class classification in hospitality review analysis, Applied Sciences (Switzerland) 14 (2024). doi:10.3390/app14146254.

[13] J. Satjathanakul, T. Siriborvornratanakul, Sentiment analysis in product reviews in thai language, International Journal of Information Technology (Singapore) (2024). doi:10.1007/s41870-024-01907-w.

[14] M. Wankhade, C. Kulkarni, A. Rao, A survey on aspect base sentiment analysis methods and challenges, Applied Soft Computing 167 (2024). doi:10.1016/j.asoc.2024.112249.

[15] R. Duma, Z. Niu, A. Nyamawe, J. Tchaye-Kondi, N. Jingili, A. Yusuf, A. Deve, Fake review detection techniques, issues, and future research directions: a literature review, Knowledge and Information Systems 66 (2024) 5071–5112. doi:10.1007/s10115-024-02118-2.

[16] Y. Mamani-Coaquira, E. Villanueva, A review on text sentiment analysis with machine learning and deep learning techniques, IEEE Access 12 (2024) 193115–193130. doi:10.1109/ACCESS.2024.3513321.

[17] M. Raza, F. Hussain, O. Hussain, M. Zhao, Z. Rehman, A comparative analysis of machine learning models for quality pillar assessment of saas services by multi-class text classification of users' reviews, Future Generation Computer Systems 101 (2019) 341–371. doi:10.1016/j.future.2019.06.022.

[18] T. Parlar, S. Özel, F. Song, Analysis of data pre-processing methods for sentiment analysis of reviews, Computer Science 20 (2019) 123–141. doi:10.7494/csci.2019.20.1.3097.

[19] G. Vinodhini, R. Chandrasekaran, A comparative performance evaluation of neural network based approach for sentiment classification of online reviews, Journal of King Saud University - Computer and Information Sciences 28 (2016) 2–12. doi:10.1016/j.jksuci.2014.03.024.

[20] K. Taha, P. Yoo, C. Yeun, D. Homouz, A. Taha, A comprehensive survey of text classification techniques and their research applications: Observational and experimental insights, Computer Science Review 54 (2024). doi:10.1016/j.cosrev.2024.100664.

[21] L. J. de Sousa, J. P. Martins, L. Sanhudo, J. S. Baptista, Automation of text document classification in the budgeting phase of the construction process: a systematic literature review, Construction Innovation 24 (2024) 292–318. doi:10.1108/CI-12-2022-0315.

[22] A. Wahdan, M. Al-Emran, K. Shaalan, A systematic review of arabic text classification: areas, applications, and future directions, Soft Computing 28 (2024) 1545–1566. doi:10.1007/s00500-023-08384-6.

[23] A. Alammary, Bert models for arabic text classification: A systematic review, Applied Sciences (Switzerland) 12 (2022). doi:10.3390/app12115720.

[24] M. Sayed, R. Salem, A. Khder, A survey of arabic text classification approaches, International Journal of Computer Applications in Technology 59 (2019) 236–251. doi:10.1504/IJCAT.2019.098601.

[25] E. Souza, D. Costa, D. Castro, D. Vitório, I. Teles, R. Almeida, T. Alves, A. Oliveira, C. Gusmão, Characterising text mining: A systematic mapping review of the portuguese language, IET Software 12 (2018) 49–75. doi:10.1049/iet-sen.2016.0226.

[26] W. Alabbas, H. Al-Khateeb, A. Mansour, Arabic text classification methods: Systematic literature review of primary studies, in: Colloquium in Information Science and Technology, CIST, Vol. 0, 2016, pp. 361–367. doi:10.1109/CIST.2016.7805072.

[27] A. Dhar, H. Mukherjee, N. Dash, K. Roy, Text categorization: past and present, Artificial Intelligence Review 54 (2021) 3007–3054. doi:10.1007/s10462-020-09919-1.

[28] C. Kesiku, A. Chaves-Villota, B. Garcia-Zapirain, Natural language processing techniques for text classification of biomedical documents: A systematic review, Information (Switzerland) 13 (2022). doi:10.3390/info13100499.

[29] M. Khadhraoui, H. Bellaaj, M. Ammar, H. Hamam, M. Jmaiel, Survey of bert-base models for scientific text classification: Covid-19 case study, Applied Sciences (Switzerland) 12 (2022). doi:10.3390/app12062891.

[30] C. Colón-Ruiz, I. Segura-Bedmar, Comparing deep learning architectures for sentiment analysis on drug reviews, Journal of Biomedical Informatics 110 (2020). doi:10.1016/j.jbi.2020.103539.

[31] G. Mujtaba, L. Shuib, N. Idris, W. Hoo, R. Raj, K. Khowaja, K. Shaikh, H. Nweke, Clinical text classification research trends: Systematic literature review and open issues, Expert Systems with Applications 116 (2019) 494–520. doi:10.1016/j.eswa.2018.09.034.

[32] G. Karystianis, K. Thayer, M. Wolfe, G. Tsafnat, Evaluation of a rule-based method for epidemiological document classification towards the automation of systematic reviews, Journal of Biomedical Informatics 70 (2017) 27–34. doi:10.1016/j.jbi.2017.04.004.

[33] M. Han, H. Wu, Z. Chen, M. Li, X. Zhang, A survey of multi-label classification based on supervised and semi-supervised learning, International Journal of Machine Learning and Cybernetics 14 (2023) 697–724. doi:10.1007/s13042-022-01658-9.

[34] K. Sikelis, G. Tsekouras, K. Kotis, Ontology-based feature selection: A survey, Future Internet 13 (2021). doi:10.3390/fi13060158.

[35] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. Acharya, V. Makarenkov, S. Nahavandi, A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Information Fusion 76 (2021) 243–297. doi:10.1016/j.inffus.2021.05.008.

[36] K. Kowsari, K. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, D. Brown, Text classification algorithms: A survey, Information (Switzerland) 10 (2019). doi:10.3390/info10040150.

[37] D. Tsirmpas, I. Gkionis, G. Papadopoulos, I. Mademlis, Neural natural language processing for long texts: A survey on classification and summarization, Engineering Applications of Artificial Intelligence 133 (2024). doi:10.1016/j.engappai.2024.108231.

[38] E. Herrewijnen, D. Nguyen, F. Bex, K. van Deemter, Human-annotated rationales and explainable text classification: a survey, Frontiers in Artificial Intelligence 7 (2024). doi:10.3389/frai.2024.1260952.

[39] M. M. M. Mironczuk, A. Muller, W. Pedrycz, The outcomes and publication standards of research descriptions in document classification: A systematic review, IEEE Access 12 (2024) 189253–189287. doi:10.1109/ACCESS.2024.3513550.

[40] O. Alyasiri, Y.-N. Cheah, H. Zhang, O. Al-Janabi, A. Abasi, Text classification based on optimization feature selection methods: a review and future directions, Multimedia Tools and Applications (2024). doi:10.1007/s11042-024-19769-6.

[41] A. Palanivinayagam, C. El-Bayeh, R. Damaševičius, Twenty years of machine-learning-based text classification: A systematic review, Algorithms 16 (2023). doi:10.3390/a16050236.

[42] A. Gasparetto, M. Marcuzzo, A. Zangari, A. Albarelli, A survey on text classification algorithms: From text to predictions, Information 13 (2022) 83. doi:10.3390/info13020083.

[43] W. Cunha, V. Mangaravite, C. Gomes, S. Canuto, E. Resende, C. Nascimento, F. Viegas, C. França, W. Martins, J. Almeida, L. Rocha, M. Gonçalves, On the cost-effectiveness of neural and non-neural approaches and representations for text classification: A comprehensive comparative study, Information Processing and Management 58 (2021). doi:10.1016/j.ipm.2020.102481.

[44] H. Wu, Y. Liu, J. Wang, Review of text classification methods on deep learning, Computers, Materials and Continua 63 (2020) 1309–1321. doi:10.32604/CMC.2020.010172.

[45] A. Kadhim, Survey on supervised machine learning techniques for automatic text classification, Artificial Intelligence Review 52 (2019) 273–292. doi:10.1007/s10462-018-09677-1.

[46] A. Shah, K. Rana, A review on supervised machine learning text categorization approaches, in: 2018 International Conference on Circuits and Systems in Digital Enterprise Technology, ICCSDET 2018, 2018. doi:10.1109/ICCSDET.2018.8821134.

[47] M. Mirończuk, J. Protasiewicz, A recent overview of the state-of-the-art elements of text classification, Expert Systems with Applications 106 (2018) 36–54. doi:10.1016/j.eswa.2018.03.058.

[48] B. Agarwal, N. Mittal, Text classification using machine learning methods-a survey, Vol. 236, 2014. doi:10.1007/978-81-322-1602-5_75.

[49] C. Aggarwal, C. Zhai, A survey of text classification algorithms, Vol. 9781461432, 2012. doi:10.1007/978-1-4614-3223-4_6.

[50] A. Bilski, A review of artificial intelligence algorithms in document classification, International Journal of Electronics and Telecommunications 57 (2011) 263–270. doi:10.2478/v10177-011-0035-6.

[51] H. Ming, W. Heyong, Filter feature selection methods for text classification: a review, Multimedia Tools and Applications 83 (2024) 2053–2091. doi:10.1007/s11042-023-15675-5.

[52] P. Pham, L. Nguyen, W. Pedrycz, B. Vo, Deep learning, graph-based text representation and classification: a survey, perspectives and challenges, Artificial Intelligence Review 56 (2023) 4893–4927. doi:10.1007/s10462-022-10265-7.

[53] Q. Li, H. Peng, J. Li, C. Xia, R. Yang, L. Sun, P. Yu, L. He, A survey on text classification: From traditional to deep learning, ACM Transactions on Intelligent Systems and Technology 13 (2022). doi:10.1145/3495162.

[54] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, J. Gao, Deep learning–based text classification, ACM Computing Surveys 54 (2022) 1–40. doi:10.1145/3439726.

[55] A. Alsaeedi, A survey of term weighting schemes for text classification, International Journal of Data Mining, Modelling and Management 12 (2020) 237–254. doi:10.1504/IJDMMM.2020.106741.

[56] X. Deng, Y. Li, J. Weng, J. Zhang, Feature selection for text classification: A review, Multimedia Tools and Applications 78 (2019) 3797–3816. doi:10.1007/s11042-018-6083-5.

[57] M. Hashemi, Web page classification: a survey of perspectives, gaps, and future directions, Multimedia Tools and Applications 79 (2020) 11921–11945. doi:10.1007/s11042-019-08373-8.

[58] A. Ullah, S. Khan, N. Nawi, Review on sentiment analysis for text classification techniques from 2010 to 2021, Multimedia Tools and Applications 82 (2023) 8137–8193. doi:10.1007/s11042-022-14112-3.

[59] E. Stamatatos, A survey of modern authorship attribution methods, Journal of the American Society for Information Science and Technology 60 (2009) 538–556. doi:10.1002/asi.21001.

[60] B. Altınel, M. Ganiz, Semantic text classification: A survey of past and recent advances, Information Processing and Management 54 (2018) 1129–1153. doi:10.1016/j.ipm.2018.08.001.

[61] G. Mujtaba, L. Shuib, R. Raj, R. Gunalan, Detection of suspicious terrorist emails using text classification: A review, Malaysian Journal of Computer Science 31 (2018) 271–299. doi:10.22452/mjcs.vol31no4.3.

[62] J. Fields, K. Chovanec, P. Madiraju, A survey of text classification with transformers: How wide? how large? how long? how accurate? how expensive? how safe?, IEEE Access 12 (2024) 6518–6531. doi:10.1109/ACCESS.2024.3349952.

[63] Y. Wu, J. Wan, A survey of text classification based on pre-trained language model, Neurocomputing 616 (2025). doi:10.1016/j.neucom.2024.128921.

[64] R. Li, M. Liu, D. Xu, J. Gao, F. Wu, L. Zhu, A Review of Machine Learning Algorithms for Text Classification, Vol. 1506 CCIS, 2022. doi:10.1007/978-981-16-9229-1_14.

[65] E. M. Guzman, V. Schlegel, R. Batista-Navarro, From outputs to insights: a survey of rationalization approaches for explainable text classification, Frontiers in Artificial Intelligence 7 (2024). doi:10.3389/frai.2024.1363531.

[66] O. Alyasiri, Y.-N. Cheah, A. Abasi, O. Al-Janabi, Wrapper and hybrid feature selection methods using metaheuristic algorithms for english text classification: A systematic review, IEEE Access 10 (2022) 39833–39852. doi:10.1109/ACCESS.2022.3165814.

[67] J. Pintas, L. Fernandes, A. Garcia, Feature selection methods for text classification: a systematic literature review, Artificial Intelligence Review 54 (2021) 6149–6200. doi:10.1007/s10462-021-09970-6.

[68] W. Weng, Y.-W. Li, J.-H. Liu, S.-X. Wu, C.-L. Chen, Multi-label classification review and opportunities, Journal of Network Intelligence 6 (2021) 255–275.

[69] N. Ur-Rahman, J. Harding, Textual data mining for industrial knowledge management and text classification: A business oriented approach, Expert Systems with Applications 39 (2012) 4729–4739. doi:10.1016/j.eswa.2011.09.124.

[70] K. Wang, Y. Ding, S. Han, Graph neural networks for text classification: a survey, Artificial Intelligence Review 57 (2024). doi:10.1007/s10462-024-10808-0.

[71] S. S. Birunda, R. K. Devi, A review on word embedding techniques for text classification, Vol. 59, 2021. doi:10.1007/978-981-15-9651-3_23.

[72] M.-A. Carbonneau, V. Cheplygina, E. Granger, G. Gagnon, Multiple instance learning: A survey of problem characteristics and applications, Pattern Recognition 77 (2018) 329–353. doi:10.1016/j.patcog.2017.10.009.

[73] S. Akpatsa, X. Li, H. Lei, A Survey and Future Perspectives of Hybrid Deep Learning Models for Text Classification, Vol. 12736 LNCS, 2021. doi:10.1007/978-3-030-78609-0_31.

[74] A. Zangari, M. Marcuzzo, M. Rizzo, L. Giudice, A. Albarelli, A. Gasparetto, Hierarchical text classification and its foundations: A review of current research, Electronics (Switzerland) 13 (2024). doi:10.3390/electronics13071199.

[75] L. da Costa, I. Oliveira, R. Fileto, Text classification using embeddings: a survey, Knowledge and Information Systems 65 (2023) 2761–2803. doi:10.1007/s10115-023-01856-z.

[76] M. Bayer, M.-A. Kaufhold, C. Reuter, A survey on data augmentation for text classification, ACM Computing Surveys 55 (2023). doi:10.1145/3544558.

[77] S. Al-shalif, N. Senan, F. Saeed, W. Ghaban, N. Ibrahim, M. Aamir, W. Sharif, A systematic literature review on meta-heuristic based feature selection techniques for text classification, PeerJ Computer Science 10 (2024) 1–45. doi:10.7717/PEERJ-CS.2084.

[78] E. Abiodun, A. Alabdulatif, O. Abiodun, M. Alawida, A. Alabdulatif, R. Alkhawaldeh, A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities, Neural Computing and Applications 33 (2021) 15091–15118. doi:10.1007/s00521-021-06406-8.

[79] J. Yang, L. Bai, Y. Guo, A survey of text classification models, in: ACM International Conference Proceeding Series, 2020, pp. 327–334. doi:10.1145/3438872.3439101.

[80] C. Ma, A. Shen, H. Yoshikawa, T. Iwakura, D. Beck, T. Baldwin, On the (in)effectiveness of images for text classification, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, 2021, pp. 42–48. doi:10.18653/v1/2021.eacl-main.4.

[81] X. Song, Y. Zhang, W. Zhang, C. He, Y. Hu, J. Wang, D. Gong, Evolutionary computation for feature selection in classification: A comprehensive survey of solutions, applications and challenges, Swarm and Evolutionary Computation 90 (2024) 101661. doi:10.1016/j.swevo.2024.101661.

[82] L. Rei, D. Mladenic, M. Dorozynski, F. Rottensteiner, T. Schleider, R. Troncy, J. S. Lozano, M. G. Salvatella, Multimodal metadata assignment for cultural heritage artifacts, Multimedia Systems 29 (2) (2022) 847–869. doi:10.1007/s00530-022-01025-2.

[83] Y. Fujinuma, S. Varia, N. Sankaran, S. Appalaraju, B. Min, Y. Vyas, A multi-modal multilingual benchmark for document image classification, in: Findings of the Association for Computational Linguistics: EMNLP 2023, Association for Computing Machinery (ACM), 2023, pp. 14361–14376. doi:10.18653/v1/2023.findings-emnlp.958.

[84] J. Hessel, L. Lee, Does my multimodal model learn cross-modal interactions? it's harder to tell than you might think!, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics (ACL), 2020, pp. 861–877. doi:10.18653/v1/2020.emnlp-main.62.

[85] Z. Chen, S. Diao, B. Wang, G. Li, X. Wan, Towards unifying medical vision-and-language pre-training via soft prompts, in: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2023, pp. 23346–23356. doi:10.1109/iccv51070.2023.02139.

[86] H. Zou, M. Shen, C. Chen, Y. Hu, D. Rajan, E. S. Chng, Unis-mmc: Multimodal classification via unimodality-supervised multimodal contrastive learning, in: Findings of the Association for Computational Linguistics: ACL 2023, Association for Computational Linguistics (ACL), 2023, pp. 659–672. doi:10.18653/v1/2023.findings-acl.41.

[87] I. Gallo, S. Nawaz, N. Landro, R. L. Grassainst, Visual Word Embedding for Text Classification, Springer International Publishing, 2021, pp. 339–352. doi:10.1007/978-3-030-68780-9_29.

[88] J. Rajendran, M. M. Khapra, S. Chandar, B. Ravindran, Bridge correlational neural networks for multilingual multimodal representation learning, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computing Machinery (ACM), 2016, pp. 171–181. doi:10.18653/v1/n16-1021.

[89] A. Ghorbanali, M. K. Sohrabi, Capsule network-based deep ensemble transfer learning for multimodal sentiment analysis, Expert Systems with Applications 239 (2024) 122454. doi:10.1016/j.eswa.2023.122454.

[90] B. Liu, L. He, Y. Xie, Y. Xiang, L. Zhu, W. Ding, Minjot: Multimodal infusion joint training for noise learning in text and multimodal classification problems, Information Fusion 102 (2024) 102071. doi:10.1016/j.inffus.2023.102071.

[91] T. Liu, Y. Hu, J. Gao, Y. Sun, B. Yin, Cross-modal multiple granularity interactive fusion network for long document classification, ACM Transactions on Knowledge Discovery from Data 18 (4) (2024) 1–24. doi:10.1145/3631711.

[92] R. Pan, J. A. García-Díaz, M. Ángel Rodríguez-García, R. Valencia-García, Spanish meacorpus 2023: A multimodal speech–text corpus for emotion analysis in spanish from natural environments, Computer Standards & Interfaces 90 (2024) 103856. doi:10.1016/j.csi.2024.103856.

[93] T. Liu, Y. Hu, J. Gao, J. Wang, Y. Sun, B. Yin, Multi-modal long document classification based on hierarchical prompt and multi-modal transformer, Neural Networks 176 (2024) 106322. doi:10.1016/j.neunet.2024.106322.

[94] Y. Li, X. Zheng, M. Zhu, J. Mei, Z. Chen, Y. Tao, Compact bilinear pooling and multi-loss network for social media multimodal classification, Signal, Image and Video Processing 18 (11) (2024) 8403–8412. doi:10.1007/s11760-024-03482-w.

[95] Y. Zhang, Image Information Prompt: Tips for Learning Large Language Models, IOS Press, 2024, pp. undefined–undefined. doi:10.3233/atde231197.
URL https://www.mendeley.com/catalogue/cb65323d-510b-31b5-a865-34cc24fff44c/

[96] S. Alqaraleh, H. Sirin, Multimodal Classifier for Disaster Response, Vol. 1983 CCIS, Springer Nature Switzerland, 2023, pp. 1–13. doi:10.1007/978-3-031-50920-9_1.

[97] D. Guo, J. Zhang, B. Yang, Y. Lin, A comparative study of speaker role identification in air traffic communication using deep learning approaches, ACM Transactions on Asian and Low-Resource Language Information Processing 22 (4) (2023) 1–17. doi:10.1145/3572792.

[98] H. Jarquín-Vásquez, D. I. Hernández-Farías, L. J. Arellano, H. J. Escalante, L. Villaseñor-Pineda, M. M. y Gómez, F. Sanchez-Vega, Overview of da-vincis at iberlef 2023: Detection of aggressive and violent incidents from social media in spanish, Procesamiento del Lenguaje Natural (2023) 351–360doi:10.26342/2023-71-27.

[99] A. Jarrahi, L. Safari, Evaluating the effectiveness of publishers' features in fake news detection on social media, Multimedia Tools and Applications 82 (2) (2022) 2913–2939. doi:10.1007/s11042-022-12668-8.

[100] Kenny, A. Chowanda, Multimodal approach for emotion recognition using feature fusion, ICIC Express Letters 17 (2023) 181–189. doi:10.24507/icicel.17.02.181.

[101] Z. Liang, Fake news detection based on multimodal inputs, Computers, Materials & Continua 75 (2) (2023) 4519–4534. doi:10.32604/cmc.2023.037035.

[102] E. Lincker, C. Guinaudeau, O. Pons, J. Dupire, C. Hudelot, V. Mousseau, I. Barbet, C. Huron, Noisy and unbalanced multimodal document classification: Textbook exercises as a use case, in: 20th International Conference on Content-based Multimedia Indexing, CBMI 2023, Association for Computing Machinery (ACM), 2023, pp. 71–78. doi:10.1145/3617233.3617239.

[103] P. H. Luz de Araujo, A. P. G. S. de Almeida, F. Ataides Braz, N. Correia da Silva, F. de Barros Vidal, T. E. de Campos, Sequence-aware multimodal page classification of brazilian legal documents, International Journal on Document Analysis and Recognition (IJDAR) 26 (1) (2022) 33–49. doi:10.1007/s10032-022-00406-7.

[104] H. Shah, K. Jaidka, L. Ungar, J. Fagan, T. Grosser, Building a multimodal classifier of email behavior: Towards a social network understanding of organizational communication, Information 14 (12) (2023) 661. doi:10.3390/info14120661.

[105] T. A. G. da Costa, R. I. Meneguette, J. Ueyama, Providing a greater precision of situational awareness of urban floods through multimodal fusion, Expert Systems with Applications 188 (2022) 115923. doi:10.1016/j.eswa.2021.115923.
URL https://linkinghub.elsevier.com/retrieve/pii/S095741742101277X

[106] L. Dongping, Y. Yingchun, S. Shikai, H. Jun, S. Haoru, Y. Qiang, H. Sunyan, D. Fei, Research on deep learning model of multimodal heterogeneous data based on lstm, IAENG International Journal of Computer Science 49 (2022).

[107] S. Kanchi, A. Pagani, H. Mokayed, M. Liwicki, D. Stricker, M. Z. Afzal, Emmdocclassifier: Efficient multimodal document image classifier for scarce data, Applied Sciences 12 (3) (2022) 1457. doi:10.3390/app12031457.

[108] O. Sapena, E. Onaindia, Multimodal classification of teaching activities from university lecture recordings, Applied Sciences 12 (9) (2022) 4785. doi:10.3390/app12094785.

[109] S. Garg, H. SS, S. Kumar, On-device document classification using multimodal features, in: Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD), CODS COMAD 2021, ACM, 2021, pp. 203–207. doi:10.1145/3430984.3431030.

[110] P. Guélorget, G. Gadek, T. Zaharia, B. Grilheres, Active learning to measure opinion and violence in french newspapers, Procedia Computer Science 192 (2021) 202–211. doi:10.1016/j.procs.2021.08.021.

[111] E. Setiawan, Multiview sentiment analysis with image-text-concept features of indonesian social media posts, International Journal of Intelligent Engineering and Systems 14 (2) (2021) 521–535. doi:10.22266/ijies2021.0430.47.

[112] X. Wang, Y. Chang, V. Sugumaran, X. Luo, P. Wang, H. Zhang, Implicit emotion relationship mining based on optimal and majority synthesis from multimodal data prediction, IEEE MultiMedia 28 (2) (2021) 96–105. doi:10.1109/mmul.2021.3071495.

[113] S. P. Zingaro, G. Lisanti, M. Gabbrielli, Multimodal side- tuning for document classification, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 5206–5213. doi:10.1109/icpr48806.2021.9413208.

[114] L. Braz, V. Teixeira, H. Pedrini, Z. Dias, ImTeNet: Image-Text Classification Network for Abnormality Detection and Automatic Reporting on Musculoskeletal Radiographs, Vol. 12558 LNBI, Springer International Publishing, 2020, pp. 150–161. doi:10.1007/978-3-030-65775-8_14.

[115] J. A. de Bruijn, H. de Moel, A. H. Weerts, M. C. de Ruiter, E. Basar, D. Eilander, J. C. J. H. Aerts, Improving the classification of flood tweets with contextual hydrological information in a multimodal neural network, Computers & Geosciences 140 (2020) 104485. doi:10.1016/j.cageo.2020.104485.

[116] W. Zhu, X. Xu, K. Yan, S. Liu, X. Yin, A synchronized word representation method with dual perceptual information, IEEE Access 8 (2020) 22335–22344. doi:10.1109/access.2020.2969983.

[117] M. Martinc, B. Škrlj, S. Pollak, Multilingual gender classification with multi-view deep learning notebook for pan at clef 2018, Vol. 2125, CEUR-WS, 2018.

[118] E. S. Tellez, S. Miranda-Jiménez, D. Moctezuma, M. Graff, V. Salgado, J. Ortiz-Bejar, Gender identification through multi-modal tweet analysis using microtc and bag of visual words: Notebook for pan at clef 2018, Vol. 2125, CEUR-WS, 2018.

[119] M. Schmitt, B. Schuller, openxbow - introducing the passau open-source crossmodal bag-of-words toolkit, Journal of Machine Learning Research 18 (2017) 1–5. arXiv:1605.06778, doi:10.48550/arxiv.1605.06778.
URL https://www.semanticscholar.org/paper/16cfa5ab8025bccca5c6bd07b62f7716d6926ade

[120] M. Cristani, C. Tomazzoli, A Multimodal Approach to Exploit Similarity in Documents, Vol. 8481, Springer International Publishing, 2014, pp. 490–499. doi:10.1007/978-3-319-07455-9_51.

[121] D. Liparas, Y. HaCohen-Kerner, A. Moumtzidou, S. Vrochidis, I. Kompatsiaris, News Articles Classification Using Random Forests and Weighted Multimodal Features, Vol. 8849, Springer International Publishing, 2014, pp. 63–75. doi:10.1007/978-3-319-12979-2_6.

[122] T. Pérez-García, C. Pérez-Sancho, J. M. Iñesta, Harmonic and instrumental information fusion for musical genre classification, in: Proceedings of 3rd international workshop on Machine learning and music, MM '10, ACM, 2010, pp. 49–52. doi:10.1145/1878003.1878020.

[123] X.-D. Zhang, Study on multi-layer fusion classification model of multi-media information, in: 2010 International Conference on Web Information Systems and Mining, Vol. 1, IEEE, 2010, pp. 216–218. doi:10.1109/wism.2010.126.

[124] S. D. Chen, V. Monga, P. Moulin, Meta-classifiers for multimodal document classification, in: 2009 IEEE International Workshop on Multimedia Signal Processing, IEEE, 2009, pp. 1–6. doi:10.1109/mmsp.2009.5293343.

[125] S. Cho, J. Moon, J. Bae, J. Kang, S. Lee, A framework for understanding unstructured financial documents using rpa and multimodal approach, Electronics 12 (4) (2023) 939. doi:10.3390/electronics12040939.

[126] D. Ortiz-Perez, P. Ruiz-Ponce, D. Tomás, J. Garcia-Rodriguez, M. F. Vizcaya-Moreno, M. Leo, A deep learning-based multimodal architecture to predict signs of dementia, Neurocomputing 548 (2023) 126413. doi:10.1016/j.neucom.2023.126413.

[127] D. Adwaith, A. K. Abishake, S. V. Raghul, E. Sivasankar, Enhancing multimodal disaster tweet classification using state-of-the-art deep learning networks, Multimedia Tools and Applications 81 (13) (2022) 18483–18501. doi:10.1007/s11042-022-12217-3.

[128] M. A. Wajid, A. Zafar, M. S. Wajid, A deep learning approach for image and text classification using neutrosophy, International Journal of Information Technology 16 (2) (2023) 853–859. doi:10.1007/s41870-023-01529-8.

[129] L. Chen, H. W. Chou, Utilizing cross-modal contrastive learning to improve item categorization bert model, in: Proceedings of The Fifth Workshop on e-Commerce and NLP (ECNLP 5), Association for Computational Linguistics (ACL), 2022, pp. 217–223. doi:10.18653/v1/2022.ecnlp-1.25.

[130] P. Kazienko, J. Bielaniewicz, M. Gruza, K. Kanclerz, K. Karanowski, P. Miłkowski, J. Kocoń, Human-centered neural reasoning for subjective content processing: Hate speech, emotions, and humor, Information Fusion 94 (2023) 43–65. doi:10.1016/j.inffus.2023.01.010.

[131] M. A. Álvarez Carmona, E. Villatoro Tello, M. Montes y Gómez, L. Villaseñor Pineda, Author profiling in social media with multimodal information, Computación y Sistemas 24 (3) (2020) 1289–1304. doi:10.13053/cys-24-3-3488.

[132] M. E. Aragón, A. P. Lopez-Monroy, A straightforward multimodal approach for author profiling: Notebook for pan at clef 2018, CEUR Workshop Proceedings 2125 (2018).
URL https://www.semanticscholar.org/paper/f728dea872696fb16e7f1f81e0c9e1c414f7e5f2

[133] D. Gupta, I. Sen, N. Sachdeva, P. Kumaraguru, A. B. Buduru, Empowering first responders through automated multimodal content moderation, in: 2018 IEEE International Conference on Cognitive Computing (ICCC), IEEE, 2018, pp. 1–8. doi:10.1109/iccc.2018.00008.

[134] Q. Gu, N. Meisinger, A.-K. Dick, Qinian at semeval-2022 task 5: Multi-modal misogyny detection and classification, in: Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), Association for Computational Linguistics (ACL), 2022, pp. 736–741. doi:10.18653/v1/2022.semeval-1.102.

[135] S. Jiang, J. Hu, C. L. Magee, J. Luo, Deep learning for technical document classification, IEEE Transactions on Engineering Management 71 (2024) 1163–1179. doi:10.1109/tem.2022.3152216.

[136] S. Bakkali, Z. Ming, M. Coustaty, M. Rusiñol, O. R. Terrades, Vlcdoc: Vision-language contrastive pre-training model for cross-modal document classification, Pattern Recognition 139 (2023) 109419. doi:10.1016/j.patcog.2023.109419.

[137] A. Rasheed, A. I. Umar, S. H. Shirazi, Z. Khan, M. Shahzad, Cover-based multiple book genre recognition using an improved multimodal network, International Journal on Document Analysis and Recognition (IJDAR) 26 (1) (2022) 65–88. doi:10.1007/s10032-022-00413-8.

[138] R. Jain, C. Wigington, Multimodal document image classification, in: 2019 International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2019, pp. 71–77. doi:10.1109/icdar.2019.00021.

[139] T. Ange, N. Roger, D. Aude, F. Claude, Semi-supervised multimodal deep learning model for polarity detection in arguments, in: 2018 International Joint Conference on Neural Networks (IJCNN), Vol. 2018-July, IEEE, 2018, pp. 1–8. doi:10.1109/ijcnn.2018.8489342.

[140] A. Chatziagapi, D. Sgouropoulos, C. Karouzos, T. Melistas, T. Giannakopoulos, A. Katsamanis, S. Narayanan, Audio and asr-based filled pause detection, in: 2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, 2022, pp. 1–7. doi:10.1109/acii55700.2022.9953889.

[141] G. Paraskevopoulos, P. Pistofidis, G. Banoutsos, E. Georgiou, V. Katsouros, Multimodal classification of safety-report observations, Applied Sciences 12 (12) (2022) 5781. doi:10.3390/app12125781.

[142] M. Ravikiran, K. Madgula, Fusing deep quick response code representations improves malware text classification, in: Proceedings of the ACM Workshop on Crossmodal Learning and Application, ICMR '19, Association for Computing Machinery (ACM), 2019, pp. 11–18. doi:10.1145/3326459.3329166.

[143] T. Yue, Y. Li, X. Shi, J. Qin, Z. Fan, Z. Hu, Papernet: A dataset and benchmark for fine-grained paper classification, Applied Sciences 12 (9) (2022) 4554. doi:10.3390/app12094554.

[144] O. Akhtiamov, V. Palkov, Gaze, Prosody and Semantics: Relevance of Various Multimodal Signals to Addressee Detection in Human-Human-Computer Conversations, Vol. 11096 LNAI, Springer International Publishing, 2018, pp. 1–10. doi:10.1007/978-3-319-99579-3_1.

[145] O. Akhtiamov, D. Ubskii, E. Feldina, A. Pugachev, A. Karpov, W. Minker, Are You Addressing Me? Multimodal Addressee Detection in Human-Human-Computer Conversations, Vol. 10458 LNAI, Springer International Publishing, 2017, pp. 152–161. doi:10.1007/978-3-319-66429-3_14.

[146] N. A. Andriyanov, Combining text and image analysis methods for solving multimodal classification problems, Pattern Recognition and Image Analysis 32 (3) (2022) 489–494. doi:10.1134/s1054661822030026.

[147] Q. Wang, T. P. Breckon, Cross-domain structure preserving projection for heterogeneous domain adaptation, Pattern Recognition 123 (2022) 108362. doi:10.1016/j.patcog.2021.108362.
URL https://linkinghub.elsevier.com/retrieve/pii/S0031320321005422

[148] U. Brefeld, Multi-view learning with dependent views, in: Proceedings of the 30th Annual ACM Symposium on Applied Computing, Vol. 13-17-April-2015 of SAC 2015, Association for Computing Machinery (ACM), 2015, pp. 865–870. doi:10.1145/2695664.2695829.

[149] M.-R. Amini, N. Usunier, C. Goutte, Learning from multiple partially observed views - an application to multilingual text categorization, Neural Information Processing Systems, 2009, pp. 28–36.
URL https://www.semanticscholar.org/paper/82402bf63b039073e2027246d1d4332781b15002

[150] Z. Jiang, T. Luo, X. Liang, Deep incomplete multi-view learning network with insufficient label information, Proceedings of the AAAI Conference on Artificial Intelligence 38 (11) (2024) 12919–12927. doi:10.1609/aaai.v38i11.29189.
URL https://ojs.aaai.org/index.php/AAAI/article/view/29189

[151] A. Doinychko, M.-R. Amini, Biconditional Generative Adversarial Networks for Multiview Learning with Missing Views, Vol. 12035 LNCS, Springer International Publishing, 2020, pp. 807–820. doi:10.1007/978-3-030-45439-5_53.

[152] A. E. Samy, Z. T. Kefato, S. Girdzijauskas, Data-Driven Self-Supervised Graph Representation Learning, Vol. 372, IOS Press, 2023, pp. 629–636. doi:10.3233/faia230325.

[153] Y. Sang, X. Mou, M. Yu, D. Wang, J. Li, J. Stanton, Mbti personality prediction for fictional characters using movie scripts, in: Findings of the Association for Computational Linguistics: EMNLP 2022, Association for Computational Linguistics (ACL), 2022, pp. 6715–6724. doi:10.18653/v1/2022.findings-emnlp.500.

[154] X. Jia, X.-Y. Jing, X. Zhu, S. Chen, B. Du, Z. Cai, Z. He, D. Yue, Semi-supervised multi-view deep discriminant representation learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (7) (2021) 2496–2509. doi:10.1109/tpami.2020.2973634.

[155] Y. Liang, H. Li, B. Guo, Z. Yu, X. Zheng, S. Samtani, D. D. Zeng, Fusion of heterogeneous attention mechanisms in multi-view convolutional neural network for text classification, Information Sciences 548 (2021) 295–312. doi:10.1016/j.ins.2020.10.021.

[156] S. Su, P. Gao, Y. Zhu, X. Liang, J. Xie, A dynamic discriminative canonical correlation analysis via adaptive weight scheme, IEEE Access 9 (2021) 142653–142663. doi:10.1109/access.2021.3118023.

[157] P. Yang, W. Gao, Information-theoretic multi-view domain adaptation: A theoretical and empirical study, Journal of Artificial Intelligence Research 49 (2014) 501–525. doi:10.1613/jair.4190.

[158] Y. Zhang, J. Wang, X. Zhang, Learning sentiment sentence representation with multiview attention model, Information Sciences 571 (2021) 459–474. doi:10.1016/j.ins.2021.05.044.

[159] F. Ma, D. Meng, X. Dong, Y. Yang, Self-paced multi-view co-training, Journal of Machine Learning Research 21 (2020).

[160] H. Wang, L. Feng, A. Kong, B. Jin, Multi-view reconstructive preserving embedding for dimension reduction, Soft Computing 24 (10) (2019) 7769–7780. doi:10.1007/s00500-019-04395-4.

[161] G. Bhatt, P. Jha, B. Raman, Representation learning using step-based deep multi-modal autoencoders, Pattern Recognition 95 (2019) 12–23. doi:10.1016/j.patcog.2019.05.032.

[162] S. Chen, L. Wang, W. Li, K. Zhang, Deep Learning Method with Attention for Extreme Multi-label Text Classification, Vol. 11672 LNAI, Springer International Publishing, 2019, pp. 179–190. doi:10.1007/978-3-030-29894-4_14.

[163] A. M. Hoyle, L. Wolf-Sonkin, H. Wallach, R. Cotterell, I. Augenstein, Combining sentiment lexica with a multi-view variational autoencoder, in: Proceedings of the 2019 Conference of the North, Vol. 1, Association for Computational Linguistics (ACL), 2019, pp. 635–640. arXiv:1904.02839, doi:10.18653/v1/n19-1065.
URL https://www.semanticscholar.org/paper/fade3b6a731f14d83bff864e87bc689c44a0399c

[164] H. Wang, J. Peng, X. Fu, Co-regularized multi-view sparse reconstruction embedding for dimension reduction, Neurocomputing 347 (2019) 191–199. doi:10.1016/j.neucom.2019.03.080.

[165] C. H. P. Ferreira, F. O. De Franca, D. R. Medeiros, Combining multiple views from a distance based feature extraction for text classification, in: 2018 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2018, pp. 1–8. doi:10.1109/cec.2018.8477772.

[166] P. Zhu, Q. Hu, Q. Hu, C. Zhang, Z. Feng, Multi-view label embedding, Pattern Recognition 84 (2018) 126–135. doi:10.1016/j.patcog.2018.07.009.

[167] Z. Zhan, Z. Ma, Document Analysis Based on Multi-view Intact Space Learning with Manifold Regularization, Vol. 10559 LNCS, Springer International Publishing, 2017, pp. 41–51. doi:10.1007/978-3-319-67777-4_4.

[168] A. Perina, N. Jojic, M. Bicego, A. Truski, Documents as multiple overlapping windows into grids of counts, Neural information processing systems foundation, 2013.
URL https://www.semanticscholar.org/paper/30db7dff9104ac512c1d15c0f48dd4c1b930abcb

[169] B. Zhang, Z.-Z. Shi, Classification of big velocity data via cross-domain canonical correlation analysis, in: 2013 IEEE International Conference on Big Data, IEEE, 2013, pp. 493–498. doi:10.1109/bigdata.2013.6691612.

[170] D. Zhang, J. He, R. Lawrence, Mi2ls: multi-instance learning from multiple informationsources, in: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Vol. Part F128815 of KDD' 13, Association for Computing Machinery (ACM), 2013, pp. 149–157. doi:10.1145/2487575.2487651.

[171] Y. Guo, M. Xiao, Cross language text classification via subspace co-regularized multi-view learning, Vol. 2, 2012, pp. 1615–1622. arXiv:1206.6481, doi:10.48550/arxiv.1206.6481.
URL https://www.semanticscholar.org/paper/f8b345f1203f153087fe23b56afad286dccb8da4

[172] M. Kovesi, C. Goutte, M.-R. Amini, Fast on-line learning for multilingual categorization, in: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, SIGIR '12, ACM, 2012, pp. 1071–1072. doi:10.1145/2348283.2348474.

[173] P. Yang, W. Gao, Q. Tan, K.-F. Wong, Information-theoretic multi-view domain adaptation, Vol. 2, 2012, pp. 270–274.
URL https://www.semanticscholar.org/paper/47b13e59e543ce671b3e0161bf89b6aa5750add3

[174] W. Zheng, Y. Qian, H. Tang, Dimensionality Reduction with Category Information Fusion and Non-negative Matrix Factorization for Text Categorization, Vol. 7004 LNAI, Springer Berlin Heidelberg, 2011, pp. 505–512. doi:10.1007/978-3-642-23896-3_62.

[175] B.-z. Zhang, W.-l. Zuo, Co-em support vector machine based text classification from positive and unlabeled examples, in: 2008 First International Conference on Intelligent Networks and Intelligent Systems, IEEE, 2008, pp. 745–748. doi:10.1109/icinis.2008.29.

[176] Z. Feng, K. Mao, H. Zhou, Adaptive micro- and macro-knowledge incorporation for hierarchical text classification, Expert Systems with Applications 248 (2024) 123374. doi:10.1016/j.eswa.2024.123374.

[177] Z. Ji, D. Kong, Y. Yang, J. Liu, Z. Li, Assl-hgat: Active semi-supervised learning empowered heterogeneous graph attention network, Knowledge-Based Systems 290 (2024) 111567. doi:10.1016/j.knosys.2024.111567.

[178] Y. Xu, G. Liu, L. Zhang, X. Shen, S. Luo, An effective multi-modal adaptive contextual feature information fusion method for chinese long text classification, Artificial Intelligence Review 57 (9) (2024) 1–29. doi:10.1007/s10462-024-10835-x.

[179] P. K. Verma, P. Agrawal, V. Madaan, R. Prodan, Mcred: multi-modal message credibility for fake news detection using bert and cnn, Journal of Ambient Intelligence and Humanized Computing 14 (8) (2022) 10617–10629. doi:10.1007/s12652-022-04338-2.

[180] F. Zhao, J. Zhang, Z. Chen, X. Zhang, Q. Xie, Topic identification of text-based expert stock comments using multi-level information fusion, Expert Systems 40 (2) (Oct. 2020). doi:10.1111/exsy.12641.

[181] C. A. Gonçalves, A. S. Vieira, C. T. Gonçalves, R. Camacho, E. L. Iglesias, L. B. Diz, A novel multi-view ensemble learning architecture to improve the structured text classification, Information 13 (6) (2022) 283. doi:10.3390/info13060283.
URL https://www.mdpi.com/2078-2489/13/6/283

[182] J. Liu, Y. Wang, L. Huang, C. Zhang, S. Zhao, Identifying adverse drug reaction-related text from social media: A multi-view active learning approach with various document representations, Information 13 (4) (2022) 189. doi:10.3390/info13040189.

[183] X. Luo, Z. Yu, Z. Zhao, W. Zhao, J.-H. Wang, Effective short text classification via the fusion of hybrid features for iot social data, Digital Communications and Networks 8 (6) (2022) 942–954. doi:10.1016/j.dcan.2022.09.015.

[184] M. Gui, X. Xu, Technology forecasting using deep learning neural network: Taking the case of robotics, IEEE Access 9 (2021) 53306–53316. doi:10.1109/access.2021.3070105.

[185] C. Hu, Y. Feng, H. Kamigaito, H. Takamura, M. Okumura, One-class text classification with multi-modal deep support vector data description, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, 2021, pp. 3378–3390. doi:10.18653/v1/2021.eacl-main.296.

[186] W. Liu, J. Pang, N. Li, X. Zhou, F. Yue, Research on multi-label text classification method based on talbert-cnn, International Journal of Computational Intelligence Systems 14 (1) (2021) 201. doi:10.1007/s44196-021-00055-4.

[187] M. M. Mirończuk, J. Protasiewicz, Recognising innovative companies by using a diversified stacked generalisation method for website classification, Applied Intelligence 50 (1) (2019) 42–60. doi:10.1007/s10489-019-01509-1.

[188] M. M. Mirończuk, J. Protasiewicz, W. Pedrycz, Empirical evaluation of feature projection algorithms for multi-view text classification, Expert Systems with Applications 130 (2019) 97–112. doi:10.1016/j.eswa.2019.04.020.

[189] J. Peng, A. J. Aved, G. Seetharaman, K. Palaniappan, Multiview boosting with information propagation for classification, IEEE Transactions on Neural Networks and Learning Systems 29 (3) (2018) 657–669. doi:10.1109/tnnls.2016.2637881.

[190] Z. Hu, Z. Zhang, H. Yang, Q. Chen, D. Zuo, A deep learning approach for predicting the quality of online health expert question-answering services, Journal of Biomedical Informatics 71 (2017) 241–253. doi:10.1016/j.jbi.2017.06.012.

[191] R. A. Sinoara, R. G. Rossi, S. O. Rezende, Semantic role-based representations in text classification, in: 2016 23rd International Conference on Pattern Recognition (ICPR), Vol. 0, IEEE, 2016, pp. 2313–2318. doi:10.1109/icpr.2016.7899981.

[192] H. Xu, M. Dong, D. Zhu, A. Kotov, A. I. Carcone, S. Naar-King, Text classification with topic-based word embedding and convolutional neural networks, in: Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '16, Association for Computing Machinery (ACM), 2016, pp. 88–97. doi:10.1145/2975167.2975176.

[193] A. Fakeri-Tabrizi, M.-R. Amini, C. Goutte, N. Usunier, Multiview self-learning, Neurocomputing 155 (2015) 117–127. doi:10.1016/j.neucom.2014.12.041.

[194] J. Liu, J. Li, X. Sun, Y. Xie, J. Lei, Q. Hu, An embedded co-adaboost based construction of software document relation coupled resource spaces for cyber–physical society, Future Generation Computer Systems 32 (2014) 198–210. doi:10.1016/j.future.2012.12.017.

[195] G. Long, J. Jiang, Graph Based Feature Augmentation for Short and Sparse Text Classification, Vol. 8346 LNAI, Springer Berlin Heidelberg, 2013, pp. 456–467. doi:10.1007/978-3-642-53914-5_39.

[196] G. Li, K. Chang, S. C. Hoi, Multiview semi-supervised learning with consensus, IEEE Transactions on Knowledge and Data Engineering 24 (11) (2012) 2040–2051. doi:10.1109/tkde.2011.160.

[197] M.-R. Amini, C. Goutte, A co-classification approach to learning from multilingual corpora, Machine Learning 79 (1–2) (2009) 105–121. doi:10.1007/s10994-009-5151-5.

[198] M. R. Amini, C. Goutte, N. Usunier, Combining coregularization and consensus-based self-training for multilingual text categorization, in: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, SIGIR '10, ACM, 2010, pp. 475–482. doi:10.1145/1835449.1835529.

[199] S. Sun, D. R. Hardoon, Active learning with extremely sparse labeled examples, Neurocomputing 73 (16–18) (2010) 2980–2988. doi:10.1016/j.neucom.2010.07.007.

[200] X.-D. Zhang, A general decision layer text classification fusion model, in: 2010 2nd International Conference on Education Technology and Computer, Vol. 5, IEEE, 2010, pp. V5–239–V5–241. doi:10.1109/icetc.2010.5529774.

[201] S. Sun, Semantic features for multi-view semi-supervised and active learning of text classification, in: 2008 IEEE International Conference on Data Mining Workshops, IEEE, 2008, pp. 731–735. doi:10.1109/icdmw.2008.13.

[202] E. T. Matsubara, M. C. Monard, G. E. Batista, Multi-view semi-supervised learning: An approach to obtain different views from text datasets, Vol. 132, IOS Press BV, 2005, pp. 97–104.

[203] V. Dasigi, R. C. Mann, V. A. Protopopescu, Information fusion for text classification — an experimental comparison, Pattern Recognition 34 (12) (2001) 2413–2425. doi:10.1016/s0031-3203(00)00171-0.

[204] M. Graff, D. Moctezuma, E. Tellez, S. Miranda, Ingeotec at da-vincis: Bag-of-words classifiers, Vol. 3496, CEUR-WS, 2023.

[205] L. Tian, X. Zhang, J. H. Lau, Metatroll: Few-shot detection of state-sponsored trolls with transformer adapters, in: Proceedings of the ACM Web Conference 2023, WWW '23, Association for Computing Machinery (ACM), 2023, pp. 1743–1753. doi:10.1145/3543507.3583417.

[206] P. Karisani, N. Karisani, L. Xiong, Multi-view active learning for short text classification in user-generated data, in: Findings of the Association for Computational Linguistics: EMNLP 2022, Association for Computational Linguistics (ACL), 2022, pp. 6441–6453. doi:10.18653/v1/2022.findings-emnlp.481.

[207] J. Li, T. Du, S. Ji, R. Zhang, Q. Lu, M. Yang, T. Wang, Textshield: Robust text classification based on multimodal embedding and neural machine translation, 2020.
URL https://www.semanticscholar.org/paper/93ad61c5700b3aabbc9192f742fea1a734bcb45b

[208] X. L. Liao, C. Zhang, A. D. Smith, G. T. Savage, A multi-topic meta-classification scheme for analyzing lobbying disclosure data, in: 2015 IEEE International Conference on Information Reuse and Integration, IEEE, 2015, pp. 349–356. doi:10.1109/iri.2015.60.

[209] B. Chen, B. Li, Z. Pan, A. Feng, Document classification with one-class multiview learning, in: 2009 International Conference on Industrial and Information Systems, IEEE, 2009, pp. 289–292. doi:10.1109/iis.2009.15.

[210] P. Gu, Q. Zhu, C. Zhang, A multi-view approach to semi-supervised document classification with incremental naive bayes, Computers & Mathematics with Applications 57 (6) (2009) 1030–1036. doi:10.1016/j.camwa.2008.10.025.

[211] X. Zhang, D.-y. Zhao, L.-w. Chen, W.-h. Min, Batch mode active learning based multi-view text classification, in: 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, Vol. 7, IEEE, 2009, pp. 472–476. doi:10.1109/fskd.2009.495.

[212] O. Akhtiamov, D. Fedotov, W. Minker, A Comparative Study of Classical and Deep Classifiers for Textual Addressee Detection in Human-Human-Machine Conversations, Vol. 11658 LNAI, Springer International Publishing, 2019, pp. 20–30. doi:10.1007/978-3-030-26061-3_3.

[213] X. Xu, W. Li, D. Xu, I. W. Tsang, Co-labeling for multi-view weakly labeled learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (6) (2016) 1113–1125. doi:10.1109/tpami.2015.2476813.

[214] X. Ma, P. Karkus, D. Hsu, W. S. Lee, Particle filter recurrent neural networks, Vol. 34, Association for the Advancement of Artificial Intelligence (AAAI), 2020, pp. 5101–5108. doi:10.1609/aaai.v34i04.5952.

[215] Y. He, Y. Tian, D. Liu, Multi-view transfer learning with privileged learning framework, Neurocomputing 335 (2019) 131–142. doi:10.1016/j.neucom.2019.01.019.

[216] C. Xu, Y. Wu, Z. Liu, Multimodal Fusion with Global and Local Features for Text Classification, Vol. 10634 LNCS, Springer International Publishing, 2017, pp. 124–134. doi:10.1007/978-3-319-70087-8_14.

[217] E. L. Iglesias, A. S. Vieira, L. B. Diz, An HMM-Based Multi-view Co-training Framework for Single-View Text Corpora, Vol. 9648, Springer International Publishing, 2016, pp. 66–78. doi:10.1007/978-3-319-32034-2_6.

[218] Y. LI, D. F. HSU, S. M. CHUNG, Combination of multiple feature selection methods for text categorization by using combinatorial fusion analysis and rank-score characteristic, International Journal on Artificial Intelligence Tools 22 (02) (2013) 1350001. doi:10.1142/s0218213013500012.