



An Enhanced Textual Review Classification and Sentiment Analysis Approach based on Machine Learning: A Comprehensive Analysis for Text Categorization Approaches

Ali Ahmed*, Nasar Ahmed, Umair Ghafoor, Syed Muhammad Rizwan, Rizwan Qureshi, Hamayun Khan, Muhammad Zunnurain Hussain

Chronicle

Article history

Received: Oct11, 2025

Received in the revised format: Nov 22, 2025

Accepted: Nov 28, 2025

Available online Dec 27, 2025

Ali Ahmed* is currently affiliated with the Faculty of Computer Science & IT Superior University Lahore, 54000, Pakistan.

Email: ali9790@gmail.com,
ali@adsultans.com

Nasar Ahmed is currently affiliated as IT Engineer, Supportiyo Ltd., Camp Hill, PA 17011, USA

Email: nasar.fazalahmed@gmail.com
nasar@supportiyo.support

Umair Ghafoor is currently affiliated as Deputy Head of Engineering Calrom Limited, M1 6EG, United Kingdom.

Email: umairghafoor@hotmail.com

Syed Muhammad Rizwan is currently affiliated with Department of Computer Engineering, University of Engineering and Technology Lahore, Pakistan.

Email: rizwan.naqvi@ieee.org

Rizwan Qureshi is currently affiliated with Department of Computer Science, COMSATS University, Islamabad, Lahore Campus Pakistan.

Email: rqureshi@cuilahore.edu.pk

Hamayun Khan is currently affiliated with the Department of Computer Science, Faculty of Computer Science & IT Superior University Lahore, 54000, Pakistan.

Email: hamayun.khan@superior.edu.pk

Muhammad Zunnurain Hussain is currently affiliated with the Bahria University Lahore Campus, Pakistan.

Email: zunnurain.bulc@bahria.edu.pk

Corresponding Author*

Keywords: Sentiment analysis, emotion recognition, sarcasm detection, explainable NLP, rationalization, transformers, and large language models, BERT, GPT, text classification.

© 2025 The Asian Academy of Business and social science research Ltd, Pakistan.

Abstract

Since interpretable feature-engineered pipelines, natural language processing (NLP) has developed into deep neural and transformer-based architectures, and finally, large language models (LLMs) that can generalize their results across tasks. This development has significantly enhanced the subjective language comprehension, such as sentiment, emotion, sarcasm, humor, stance, metaphor, intent and aesthetic judgments, and increased the requirements of explainability in high-stakes areas. Already trained transformers (e.g., encoder-only and decoder-only versions) and LLMs like BERT and GPT have established powerful baselines on text classification and zero/few-shot subjectivity modeling, but the opaqueness of these models drives language-based rationalization (extractive/abstractive natural-language justifications) and feature-attribution algorithms (e.g., LIME, Integrated Gradients, SHAP). The four strands brought together by this survey include: (i) sentiment-analysis methods and datasets (e.g., IMDb, Sentiment140, Twitter Airline, SemEval), (ii) transformer-era text classification, (iii) LLMs on subjective language, and (iv) rationalization of explainable NLP. We have provided a unified taxonomy of subjective tasks and have analyzed model families, starting with classical ML up to transformers/LLMs; have collected major data sets and benchmarks; and have systematized explainability methods, in particular rationalization, and their evaluations and classify open problems dataset bias and annotation ambiguity, constraints of faithfulness and explanatory comprehensibility, evaluation bias, compute cost, and ethical risks. The article will facilitate the formation of a coherent foundation of explainable, credible subjective NLP based on the functions of transformer/LLM and analyze feature extraction techniques methodically based on LLM. We have trained a Machine Learning classifier using 70% of the training data and 30% of the testing data. Based on our results, we find that the proposed ML based technique gives enhanced performance, with an improved accuracy of 99% in the UCI-ML reviews dataset and 96% in the Twitter Kaggle dataset. This study underscores the paramount significance of feature extraction in sentiment analysis, endowing pragmatic insights to elevate model performance and steer future research pursuits.

INTRODUCTION

Sentiment analysis is known as a method in which a statement is divided into small parts, learning from them (like what these parts perform) that how they communicate to one another. Analysis of human behavior according to their textual contents is called Opinion Mining (OM). Online consumption has become more popular with the fast penetration of the Internet into the everyday lives of people. New forms of communication between customers and service providers can be offered by its emergence and pervasiveness. Consumers are likely to use e-commerce websites to share their experiences of consumption and voice their views on any product or service, after the consumption behavior. The amount of online reviews created on review sites, forums, blogs etc., can be viewed as a big data challenge to the consumers and organizations themselves, should they fail to analyze and make the best use of them. The challenge of this data deluge can be tackled with cognitive computing by this means of processing this multimodal data, besides transforming raw data into useful information [1, 2]. The focus on cognitive computing has increased recently due to its capacity to offer an interdisciplinary approach and a set of technologies to analyze big data, attaining intelligent cognitive processes by cutting through the complexity of big data that would help enhance human decision-making [3].

Early sentiment mining and opinion mining pipelines employed transparent feature engineering (e.g., TF-IDF, n-grams, lexicons) in linear/ probabilistic classifiers (Naive Bayes, SVM, logistic regression). These techniques were interpretable with explicit feature weights and decision-making rules that the practitioner could use to interpret the predictions as being made by a particular linguistic pattern. Nevertheless, they could not make broad cross-domain and fine-grained context generalization due to their dependence on hand-crafted characteristics. Neural models (CNNs, RNNs, LSTMs) were more accurate and less interpretable. Convolutional architectures learned filters representing local n-gram statistics, and recurrent architectures learned sequential dependencies and long-range context. The developments facilitated improved action on intricate subjective problems, yet the acquired representations grew more opaque, and it was hard to describe individual forecasts [4, 5]. The transformer architecture and pretraining paradigm established a general-purpose framework for text classification and subjective tasks. BERT introduced bidirectional contextualized representations through masked language modeling, while GPT demonstrated the power of autoregressive language modeling. These architectures became de facto backbones for downstream classification and inference tasks [6]. The self-attention mechanism enabled modeling of long-range dependencies without the vanishing gradient problems of RNNs, leading to substantial performance improvements across benchmarks [7].

$$g^t(x) = \sum_{i=1}^t y_i \prod_{j=1, j \neq i}^t \frac{(x - x_j)}{(x_i - x_j)}$$

Eq (1)

Massive scale, tuned to instructions, and few-shot learning Large language models (LLMs) like GPT-3 [8], PaLM [9], LLaMA [10], and GPT-4 [11] extended capabilities by large scale, instruction tuning, and few-shot learning. The remarkable zero-shot and

few-shot performance on subjective tasks without task-specific fine-tuning seen in these models is fundamentally changing our approach to NLP applications.

$$S = a_0 = g(0) = \sum_{i=1}^t g(i) \prod_{j=1, j \neq i}^t \frac{-j}{(i-j)} \pmod{p}$$

Eq (2)

Simultaneously, explainable NLP evolved to feature-attribution and saliency-type tools (e.g., LIME [12], Integrated Gradients [13], SHAP [14]) to rationalization, where the explanations are expressed in natural language through extracting the spans of evidence or creating justifications [15]. Influential input tokens are emphasized by feature-attribution techniques, which are gradient-based or perturbation-based, yet necessitate technical skills to interpret and do not have natural-language explanations [15]. Rationalization handles these lacks by generating human-understandable textual explanations that can be comprehended by non-expert users [16].

$$G_B = \sum_{i=1}^N W_i \cdot G_{L_i}$$

Eq (3)

Cognitive computing

Cognitive computing is a relatively new research area that derives its inspiration in the intersection of cognitive science, data science, and computing technologies [17]. The cognitive computing definition is a controversial subject matter. Academic and industrial ranges of thought have described and defined cognitive computing in a variety of ways. Through the applications and service perspective, cognitive computing may offer a set of processes that allow computer systems to extend and multiply the human expertise by replicating intelligent behaviors of the human mind in reasoning, remembering and learning through experience to perform a wide range of cognitive functions. An analogy of human beings can be useful in getting to know the role of cognitive computing and its mechanism of working. As a study field of the cognitive science profession, cognition has been defined as the mental powers or functions of learning and knowing using senses, thinking, and experience. In the case of human cognition, the cognitive processes use organismal means of their realization i.e. the brain and the mind [18].

$$Q^{(i)} = \{Q(x_j^{(i)})\}_{j=1}^{m_i}$$

Eq (4)

The machine cognition is to cognitive computing like the brain or mind to human cognition. Implementation of cognitive computing depends on cognitive analytics being offered by data science and computing technologies offered by computing disciplines.

$$fd_k(x) = \frac{1}{n} \sum_i^{n_m} l(p_i, q_i; x)$$

Eq (5)

The cognitive computing systems are allowed to perform cognitive tasks of observation, interpretation, assessment, and decision, which is similar to human brain work, with the help of cognitive technologies, including big data, cloud computing, artificial intelligence, machine learning, pattern recognition, data mining and natural language processing. In total, cognitive computing draws various interdisciplinary theories, techniques, and tools to simulate human cognition, driving cognitive systems that can perceive, reason, learn, and behave independently and hence providing real human-machine cooperation [19].

$$\text{minimum } fd_k(x) = \sum_{m=1}^{C * M} \frac{1}{n} fd_m(x)$$

Eq (6)

Two main types of rationalization are identified, extractive rationalization which chooses critical terms or phrases of the input as the rationale, and abstractive rationalization which creates a new explanatory text that summarizes the reasoning. Current LLMs support either of these forms by enabling them to make natural-language generation and prompt them to give a series of reasons with chain-of-thought methods. The non-expert audiences find rationalization more appealing since it is employed to convey reasoning in written form, as opposed to numerical attributions, which makes AI systems more understandable and acceptable. This is especially crucial in subjective tasks where context, tone as well as implicit meaning are greatly important in human perception [21].

The result of sentiment classification, sentiment orientation has been a major reference as well as an influencer not only to individuals but also to organizations in the decision-making process. The sentiment polarity present in the user reviews has been observed to be among the most effective factors that contribute to the usefulness of the user reviews in the aspect of making a purchase decision and consumers have the instinct of making a purchase of a product that has positive reviews. Conversely, depending on the sentiment orientation of the consumers' reviews, the organization can formulate its marketing strategy, including amending its brand positioning, launching marketing messages, designing new products and other operations to enhance the organizational performance. Consequently, the quality of the sentiment classification directly indicates the possibility of consumers and organizations to make smart decisions or not, depending on reviews and poor results may result in ineffective or even incorrect decisions.

ELU – E- Linear Unit with $0 < \alpha$ is

$$f(x) = \begin{cases} \alpha(\exp(x) - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

Eq (7)

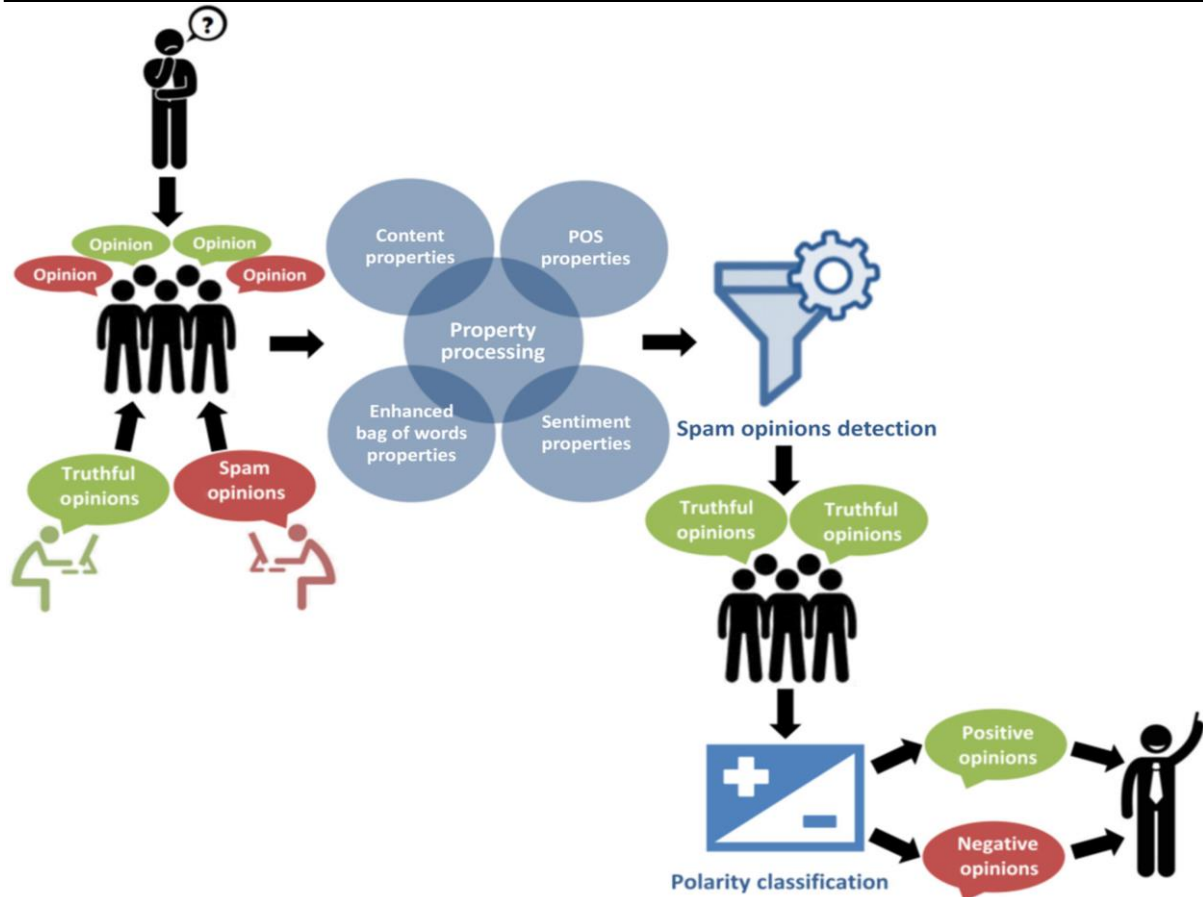


Figure 1.
Generalize enhanced Sentiment Analysis approach [22]
Natural language processing (NLP)

Natural language processing (NLP) is a very crucial aspect in the area of cognitive computing, as it provides a cycling learning process such that cognitive systems may be used to make sound data-driven decisions. The ability to process unstructured data within a context where questions are posed and to give intelligent cognitive responses, in other words, to enable machines to communicate with humans is one of the most basic tasks in cognitive computing processes. This is what NLP does best, i.e., to convert an unstructured material into a cognitive knowledge base that determines the meaning of natural language [23]. NLP relies on statistical applications, language rules and other methods like machine learning.

These tools may be used to perform various linguistic analyses, i.e. lexical analysis, syntactical analysis, discourse analysis etc. NLP, therefore, can automate the thinking process of deriving the meaning of words, recognition of patterns and association between words, and establishment of layers of context perception. Among the NLP-oriented applications that have been introduced in the field of cognitive computing is the ability to comprehend the word-of-mouth (WOM) of the text-based remarks posted by customers in the online web, with the help of sentiment analysis to find out the actual feelings or attitudes of the customers towards any product or service. Sentiment analysis, sentiment classification, or opinion mining is a binary classification task of categorizing the textual reviews in terms of being either positive or negative polarity [24, 25].

Motivation

Transformer and LLM systems achieve state-of-the-art results but behave as opaque black boxes, making it difficult to understand *why* a subjective prediction was produced (e.g., irony vs. literal, positive vs. negative). In sensitive domains such as healthcare, legal decision-making, content moderation, and financial analysis, opacity jeopardizes trust, auditability, and safety. Users and stakeholders need to understand not just *what* the model predicted, but *why* it made that prediction, to assess reliability and identify potential biases or errors [26].

In the meantime, subjective tasks tend to be ambiguous and situation-specific (sarcasm, humor, aesthetics), typically resulting in discrepancies between the annotators and difficulty in evaluation. Detection of sarcasm, such as, can be based on mutual cultural context, tone of voice or situational knowledge not found in the text itself [27]. The humor recognition is about the incongruity, surprise and subjective appreciation, which is different to individuals and cultures. Emotion recognition has to negotiate the multidimensional affective conditions, which can include a mixture of different emotions or represent the expression of culturally-specific emotional categories [28]. Such concurrent demands of high performance and low transparency drive a single approach to subjectivity and elucidate the ability of LLMs to make fine judgments and the rationalization process to reveal the reasoning behind the judgment in an accurate way [29, 30]. It is required that the field should offer predictively accurate techniques, offer meaningful explanations that make users trust and appropriate to rely on model outputs [31].

This work incorporates four quality sources and the foundations used in citing them: Sentiment analysis approaches and datasets: including classical approaches to machine learning, deep learning, and transformers, with standard datasets such as IMDb, Sentiment140, Twitter US Airline, and SemEval shared tasks [32, 33]. Transformer-based text classification: architectural overview and overview of the BERT-family encoders, GPT-family decoders, and their classification task adaptations, efficiency and safety concerns [34, 35]. LLMs for subjective language: extensive taxonomy of sentiment, emotion, sarcasm, humor, stance, metaphor, intent, and aesthetics with evaluation of zero-shot and few-shot performance [36, 37].

Table 1.
Analysis of Textual data Model

Ref	Model	Task	Performance
[38]	Dynamic convolutional neural network (DCNN)	Binary sentiment prediction	The accuracy of binary sentiment classification is 86.8%
[39]	Deep recursive neural network (DRNN)	Multi-class sentiment prediction	The accuracy of fine-grained sentiment classification is 48.5%
[40]	Uniform-layer LSTM architecture	Twitter sentiment prediction	The accuracy of six-way question classification is 93.0%
[41]	Uniform-layer LSTM architecture	Sentence-level sentiment classification	Uniform-layer model achieves average 0.8%
[42]	Coupled-layer LSTM architecture	classification	improvement of the performances
[43]	Shared-layer LSTM architecture	Document-level sentiment classification	Coupled-layer model achieves

Explainable NLP via rationalization, extractive and abstractive methods, schemes of evaluation based on fidelity and plausibility, and human-oriented evaluation methods. The goal is to standardize task taxonomies, model families,

datasets/benchmarks, and explainability approaches, rigorously under the mentioned materials, to be useful in helping trace scholarly research and high-density IEEE referencing. This combination will give the researchers and practitioners a whole picture and learn more about the actual situation in the field of subjective NLP and explainability, including gaps and how to further research. Coherent vision of subjectivity + explainability. We unify sentiment analysis (to interpretable ML to DL to transformers/LLMs), subjective task taxonomies, and explainability systems into a unified system, rationalizing it to be a first-class subjective NLP component [44, 45]. The integration points out the natural synergy between subjective language comprehension, which involves the necessity of both contextual and practical reasoning and the explanation generation that also necessitates these forms of reasoning abilities. Unified taxonomic classification of subjective tasks. We are using an eight-task perspective sentiment, emotion, sarcasm, humor, stance, metaphor, intent, aesthetics, which aligns with current surveys in the era of the LLM [46-49]. This taxonomy offers an orderly framework of how research can be organized, what similarities can be found in tasks, and what challenges are task-generated. Comparison of paradigm families.

We are tracking the development of interpretable ML (NB/SVM/LR with TF-IDF/n-grams) to deep learning (CNN/LSTM/GRU) and transformers/LLMs (BERT/GPT and successors), and its implications to subjective NLP [50, 51] Every paradigm shift led to the increase in performance as well as restructured the interpretability landscape which required the novel methods of explanation. Datasets and benchmarks. We select a task-wise and dataset-wise view of reviews and social media: IMDb (50k balanced), Sentiment140 (1.6M tweets), Twitter US Airline (14k, 3-class), SemEval-2017 Task 4 (Twitter subtasks), GoEmotions (58k Reddit comments, 27 emotions), SARC (sarcasm corpus), and numerous others [52, 53]. This detailed data catalog assists scientists in choosing the right standards and knowing what domains are represented and what the constraints are.

Explainability and Datasets Involved

evaluation and rationalization methods. We systematize extractive vs. abstractive rationalization [54] detail common attribution baselines (LIME [55], Integrated Gradients [56], SHAP [57]), and summarize benchmarking datasets like ERASER [58] and e-SNLI [59]] with human-centered evaluations [60]. We distinguish between fidelity (faithfulness to model reasoning) and plausibility (human-judged quality), guiding evaluation methodology. Open issues and directions. We consolidate reported limitations in annotation quality and ambiguity, dataset bias, faithfulness/comprehensibility gaps in explanations, computational cost, and the need for standardized evaluation [61, 52].

$$\delta_h = 60^\circ \begin{cases} 0 + \frac{(\beta_g - \beta_b)}{(m_x - m_n)}, \text{ if } m_x = \beta_r \\ 2 + \frac{(\beta_b - \beta_r)}{(m_x - m_n)}, \text{ if } m_x = \beta_g \\ 4 + \frac{(\beta_r - \beta_g)}{(m_x - m_n)}, \text{ if } m_x = \beta_b \end{cases} \quad \text{Eq (8)}$$

Table 2.
Major Datasets for Sentiment Analysis

Dataset	Size	Domain	Classes	Ref
IMDb Movie Reviews	50,000	Movies	Binary (pos/neg)	[53]
Sentiment140	1.6M	Twitter	Binary (pos/neg)	[54]

Dataset	Size	Domain	Classes	Ref
Twitter US Airline	~14,000	Customer service	3-class (pos/neu/neg)	[55]
SemEval-2017 Task 4	Varied	Twitter	Multi-task	[56]
Amazon Product Reviews	Millions	E-commerce	5-star scale	[57]
Yelp Reviews	1.6M+	Restaurant/business	5-star scale	[58]

Emotion Recognition

Definition: Automatic recognition of the affective or emotional expression of text: e.g., joy, anger, sadness, fear, surprise, disgust, or continuous dimensions (valence, arousal, dominance) [59]. Affect ambiguity and co-occurrence (mixed emotions), Cultural diversity in emotion expression, Class imbalance (joy prevails in many corpora), Fine-grained emotions (fear and anxiety) distinction.

Table 3.

Major Datasets for Emotion Recognition

Dataset	Size	Emotions	Domain	Ref
ISEAR	~7,600	7 basic	Surveys	[60]
GoEmotions	58,000	27 fine-grained	Reddit	[61]
SemEval-2018 Task 1	~10,000	11 emotions	Twitter	[62]
DailyDialog	13,000	7 basic	Dialog	[63]
MELD	13,000	7 basic	Multimodal TV	[64]

Overall, sentiment analysis is done in two types of techniques. One of them is machine learning-based methods and the other is lexicon-based methods [65]. The latter is disadvantaged by the fact that machine learning-based methods have a high reputation in academia and in the industrial world because their performance is often better as compared to the former. Deep learning is a form of machine learning algorithm and it has been successful due to the increased capabilities of chip processors, reduced hardware prices and it has also led to the improvement of machine learning algorithms [66].

$$\delta_s = \left(\frac{m_x - m_n}{m_n} \right) \quad \text{Eq (9)}$$

Precise sentiment analysis in e-commerce, multiple research works suggest new machine learning methods or enhance the current algorithms, suggesting to report a greater degree of accuracy because of the application of potent algorithms to the feature extraction [67] or classifier design [68].

$$\delta_v = m_x(\beta_r, \beta_g \beta_b), \delta_{sv} = m_n(\beta_r, \beta_g \beta_b) \quad \text{Eq (10)}$$

These problems are important research challenges that can be used in the further development of the field. The rest of this survey is structured in the following manner: Section 2 formalizes foundations and definitions of subjectivity and explainability, defining the major concepts and terms. Section 3 outlines our survey methodology and literature search strategy as well as the taxonomy formation process. Section 4 shows the single taxonomy of subjective tasks, which includes sentiment, emotion, sarcasm, humor, stance, metaphor, intent, and aesthetics. Section 5 reviews algorithms and models from classical ML through transformers/LLMs, analyzing the evolution of approaches and their implications for explainability. Section 6 includes datasets and benchmarks on tasks, which is a comprehensive contribution to researchers. Section 7 details explainability and rationalization techniques, evaluation frameworks, and integration with modern architectures. Section 8 offers a

comparative integration of tasks, models, as well as methods of explanation. Section 9. The issues and unresolved problems of the subjective NLP and explainability are discussed here. Section 10 describes the future research directions and Section 11 the final section of the survey, is the conclusions and recommendations.

LITERATURE REVIEW

Foundations of Subjective and Explainable NLP

The term subjectivity in natural language is used to describe the articulation of individual views, judgments, beliefs, feelings or feelings of affection as opposed to objective and verifiable facts. According to linguistic research, subjectivity can be defined as the encoding of personal stance, which is expressed by opinion adjectives, intensifiers, first-person references, hedges, or modality cues [69].

Computationally, subjectivity classification draws the line between those sentences that have an opinion, sentiment, or emotion and those that are factual. Subjective language, which uses subjective language, is cognitively based on the internal state or purpose of a speaker, and is based on Theory of Mind reasoning and pragmatic inference. Such language may only be comprehended in terms of context and culture since the text of such language is never understood literally, as, e.g., sarcasm (Great, another meeting! meaning frustration) and metaphor (Time is money meaning abstract relationships) can be understood in terms of their literal meaning and intended meaning, respectively [70].

$$R(t) = \sum_{i=1}^n FI_i(t) * \tau_{ih} + [\tau_h * r_{i-1}] \quad \text{Eq (11)}$$

The first computational methods of subjectivity were lexicon-based (e.g., MPQA, SentiWordNet) and rule-of-thumb-based methods to detect opinionated language [71]. Subsequent neural and transformer architectures were able to pick up finer contextual details in learned representations, allowing implicit sentiment and figurative language to be done in a better way [72, 73].

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad \text{Eq (12)}$$

NLP Explainability and Rationalization

Explainability aims to render model predictions clear and understandable to allow users to know how decisions were made [74, 75]. Due to the high performance of deep models as well as transformer-based models, which are black-box.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad \text{Eq (13)}$$

LIME (Local Interpretable Model-agnostic Explanations) [76] generates local linear model approximations around particular model predictions. Integrated Gradients [77] predicts the score of an attribute using the integration of the gradient on the input features. SHAP (Shapley Additive exPlanations) [78] Shapley values, which are computed using the game-theoretic model, are used to assign importance to features. Such techniques emphasize influential input tokens, but demand technical

skills and do not tend to be naturally language justified. Rationalization is a way to fill these holes through the generation of human-readable textual explanations [79]. Two major forms are identified. Chooses the least, adequate input spans (words/sentences) to make a prediction. The rationale is a part of the original input text that is used to justify the decision of the model. Attention-based selection based on learned attention weights, learning by reinforcement to determine informative spans, and Differentiable sampling with hard selection.

$$TDI = \sqrt{(\Delta C)^2 + (\Delta \sigma)^2} \tag{Eq (14)}$$

Abstractive Rationalization and Stance Detection

Produces free textual explanations of the reasoning. The strategy involves the application of sequence-to-sequence models or the LLM decoders to generate new explanations that are not limited to the input text. Examples include: Explanations in e-SNLI using natural languages [80], Chain-of-thought reasoning in LLMs [81], Predictions are generated with justifications [82]. The modern versions of LLMs support both types due to the natural-language generation features, and they can be asked to explain step-by-step [6][134]. Rationalization offers the local explanations to individual predictions and is complementary to the traditional feature-attribution approaches since it can disclose the linguistic evidence that leads to a model output [83, 84].

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))}} \tag{Eq (15)}$$

Table 4. Major Datasets for Sarcasm Detection

	Dataset	Size	Modality	Domain	Reference
Sarcasm Detection	SARC (Reddit)	1.3M	Text	Social media	[85]
	Twitter Sarcasm	~100k	Text	Twitter	[86]
	iSarcasmEval 2022	~5,000	Text	Twitter/Reddit	[87]
	MUSTARD	690	Multimodal	TV shows	[88]
	MUSTARD++	345	Multimodal	TV shows	[89]
	Pun of the Day	~2,400	Puns	Various	[90]
Humor Detection	Humicroedit	~15,000	Micro-edits	News headlines	[91]
	SemEval-2021 Task 7	~10,000	Rating & binary	Reddit/Twitter	[92]
	Short Jokes Dataset	~200,000	Short jokes	Reddit	[93]
	Pun of the Day	~2,400	Puns	Various	[94]

Contrast words, punctuation (!), sentiment flip detection & Neural: BiLSTM/CNN with attention to context [95], Transformers: RoBERTa with discourse context [96], Multimodal: Audio-visual cues combined with text, LLMs: Few-shot prompting with explanation generation [97, 98] Determining the presence of humor, wit or comedic intent in text [99, 100]. Humor often involves discrepancy, surprise, pun or contextual reversal. Very subjective (different people have different preferences), Cultural peculiarity (jokes do not) never translate), Type of humor (puns, satire, dark humor), and Scarcity of training data. The identification of whether text is favorable, unfavorable or neutral in its position concerning a. specific target, issue, or claim [101, 102]. As opposed to sentiment, stance is target-dependent. The Major Problems are Target identification and disambiguation, Implicit stance (no explicit opinion words), multi-target situations, Fact-checking and rumor detection applications.

$$Y(t) = \omega[\tau_{ho} * h(t)]$$

Eq (16)

Stance and Metaphor Recognition

Target-based feature engineering & Neural: Attention mechanisms between text and target [103], Transformers: BERT, it has target encoding [104], LLMs: Instruction target mention on tuned models [105, 106]. Finding figurative expressions where a word or phrase is applied symbolically instead of. Literally [107, 108]. Example: her voice is music to my ears. The voice is analogized to music. The Key Challenges are Traditional and novel metaphors, Cross-domain mappings (abstract ← concrete), differentiating literal and figurative usage, Limited annotated corpus

$$\omega = E_f * \frac{1}{1 + e^{-\theta t_f}}$$

Eq (17)

Table 5.
Major Datasets for Stance Detection

	Dataset	Size	Targets	Domain	Reference
Stance Detection	SemEval-2016 Task 6	~4,000	6 topics	Twitter	[107]
	RumourEval 2017/2019	~8,000	Rumor threads	Twitter	[108]
	FakeNewsNet	~20,000	News articles	News/social	[109]
	VU Amsterdam	~200,000	All-POS	Twitter	[110]
	Metaphor Corpus	tokens			
Metaphor Recognition	MOH-X	~15,000	Verb metaphors	Twitter	[111]
	TroFi	~3,700	Verb usage	Twitter	[112]

Relationship Between Subjectivity, Sentiment, and Explainability

Subjectivity, sentiment, and explainability are related, with subjectivity being the least explainable or the most imprecise, and sentiment the most imprecise and explainable, which is understandable. The relationship between subjectivity, sentiment, and explainability is, loosely, as follows: the less explainable, the more imprecise the subjectivity, and the more imprecise yet explainable, sentiment. The intersection between subjectivity and explainability lies in both focusing on interpretation [113, 114]. The underlying subjective task sentiment analysis originally relied on decipherable lexicons and feature weights that revealed decision logic [115, 116]. The Bag-of-words and TF-IDF representations enabled practitioners to view the words that were associated with positive or negative labels, which inherently gave transparency [117, 118]. With the increase in the complexity of architectures (CNNs, LSTMs, transformers), interpretability decreased since features were now latent and spread across high-dimensional representations [119, 120]. This obscurity spurred rationalization and attention-visualization techniques in order to regain transparency [121, 122]. Explainability gives subjective NLP two important enhancements:

Clarifying ambiguity

Differentiating between sarcasm and actual praise, detecting emotion stimuli, or disclosing metaphorical mappings by referring to contextual clues [123, 124]. By enabling users to check the validity of textual evidence in supporting model predictions, especially in affective computing systems used in mental health, customer service, and content moderation, this can be encouraged [125, 126].

Therefore, rationalization is not merely peripheral but integral to modeling human-like subjectivity [127].

Deep learning for sentiment analysis

Deep learning or deep neural networks (DNNs) is a sub-field of machine learning algorithms that has been broadly used in classic areas of artificial intelligence, including computer vision [128], speech recognition and natural language processing [129]. Deep learning can reach the state of the art when applied to sentiment analysis, which is also one of the most popular issues in NLP, since it can learn representations of data with many layers of abstraction by using computational models made of many processing layers [130].

$$\delta_h = 60^\circ \begin{cases} 0 + \frac{(\beta_g - \beta_b)}{(m_x - m_n)}, \text{ if } m_x = \beta_r \\ 2 + \frac{(\beta_b - \beta_r)}{(m_x - m_n)}, \text{ if } m_x = \beta_g \\ 4 + \frac{(\beta_r - \beta_g)}{(m_x - m_n)}, \text{ if } m_x = \beta_b \end{cases} \quad \text{Eq (18)}$$

Due to this, sentiment analysis is increasingly being conducted by researchers using deep learning methods. Different network architectures that are effective in deep learning are the convolutional neural network (CNN), recurrent neural network (RNN), recursive neural network (RNN), deep belief network (DBN), among others. As a rule, the existing study on the topic takes two directions. Other studies focus on the construction of powerful network structures based on simple deep learning models to address various problems in sentiment analysis tasks [131], whereas others tend to make comparative analyses of the various basic deep neural networks on a given task to offer practical advice in the choice of deep learning models [132, 133]. This section briefly outlines some of the representative research methodologies on sentiment analysis with the use of deep learning methods. The dynamic convolutional neural network (DCNN) is a CNN network that is used to semantically model a sentence in when it comes to the task of sentiment analysis. The network considers the input sentences of different lengths and explicitly captures word relations of different sizes by constructing a feature graph on the sentence. It also happened that the network is very efficient in a range of sentiment analysis tasks. In [134] suggested three different information sharing architectures were suggested with the RNN architecture i.e., uniform-layer architecture, coupled-layer architecture, and shared-layer architecture.

$$\delta_s = \left(\frac{m_x - m_n}{m_n} \right) \quad \text{Eq (19)}$$

They incorporated an RNN within a multi-learning framework to project arbitrary text directly to semantic representations in the form of vectors with task-specific and shared layers and they have shown that their models may be used to boost the performance of a classification task in the presence of other related tasks. To enable the possibility of the capacity of a hierarchical representation, [135] proposed a deep recursive neural network (deep RNN) using a stack of recursive layers. They evaluated the network through the fine-grained sentiment classification and the outcome revealed that their method did better than earlier baselines in the sentiment analysis task. The semi-supervised recursive autoencoders (RAE) in [136] learns to learn the

representations of the multi-word phrases in the form of vectors. The model is able to perform sentiment prediction tasks better than other state-of-the-art algorithms, and predict sentence-level sentiment distribution better than a number of baselines. [137] introduced a VBN framework of learning word dependencies in text, which has a constant number of variables. They trained dynamic Gaussian Bayesian networks to get the starting weights of the neurons in the hidden layer, and trained the time-delayed network in a deep DBN network.

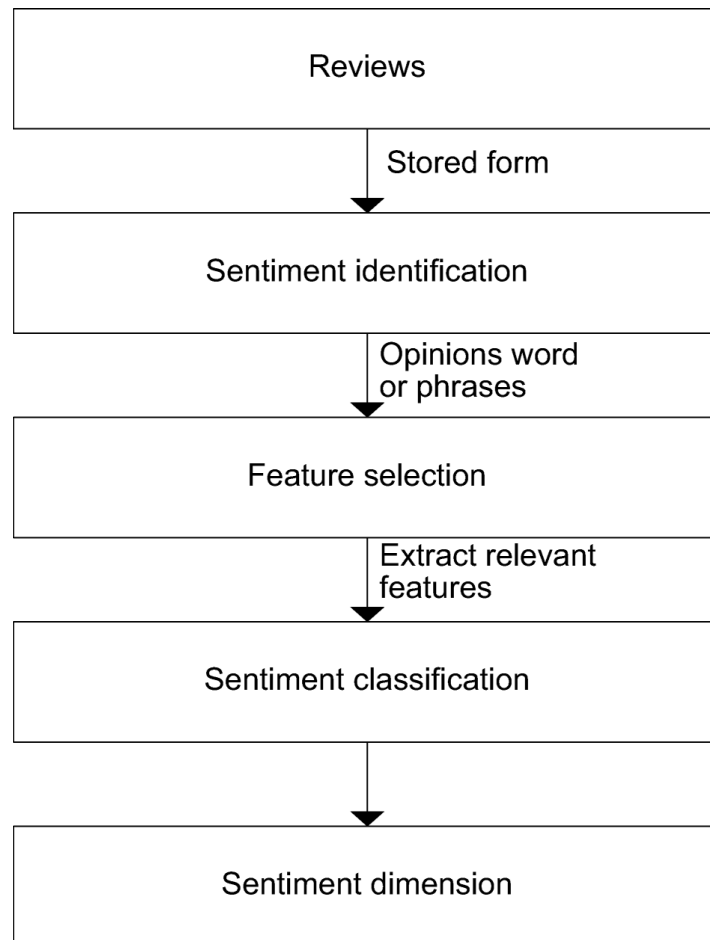


Figure 2.
Generalize Process of sentiment analysis Approach [138]

$$\delta_v = m_x(\beta_r, \beta_g \beta_b,), \delta_{sv} = m_n(\beta_r, \beta_g \beta_b,) \tag{Eq (20)}$$

The experimental findings showed that VBN can make more than 30% improvement over state-of-the-art baselines. Studied four dissimilar kinds of features, which are linguistic category characteristics, review metadata, readability and subjectivity, which lead to the quality of online reviews, and suggested a new predictive model of review helpfulness by utilizing them. The prediction model was created using three machine learning algorithms (Naïve Bayes, Support Vector Machine and random forest). It was discovered that the proposed model was able to provide a predictive precision of more than 77 percent on actual review data, and a blend of attributes gives the best predictive score. It was also clear that these aspects might be powerful predictors of the quality of the review. The reviewer-related characteristics of reviews are also examined in other studies to compare the usefulness of reviews [139].

$$R(t) = \sum_{i=1}^n FI_i(t) * \tau_{ih} + [\tau_h * r_{i-1}] \quad \text{Eq (21)}$$

Examined not only a quantitative aspect (i.e., message length) but also qualitative aspects (e.g., reviewer experience, reviewer impact, product rating) with their influence on the helpfulness of the review. To test their hypotheses, a Tobit regression of review data was run on the data of Amazon.com. They confirmed that the word count is a significant predictor of the helpfulness [140, 141], however, they observed that the positive influence of the word count on the helpfulness might decrease considerably or even reverse at some point when the word count surpasses some threshold. In particular, their empirical findings showed that the threshold may be as small as 144 words. As to the qualitative factors, the experience of the reviewer and the influence of the reviewer were reported to have no significance on the predictors of helpfulness, but the review framing was proven to be effective and contributed to the online review helpfulness, as the research conducted by [142].

$$Y(t) = \omega[\tau_{ho} * h(t)] \quad \text{Eq (22)}$$

The work by [143] was different compared to other studies that applied the signaling theory in order to create an advanced model of the investigation of the variables that impact the quality and usefulness of the review. Regarding the theoretical backgrounds of signaling theory, they singled out two classes of signals, namely, content-related signals of reviews (particular content and writing styles) and reviewer-related signals (reviewer knowledge and non-anonymity), to attempt to model interrelations between them and the review helpfulness. Using the proposed model, the analysis of the Tobit regression revealed that the drivers of the higher quality review were a greater number of signals linked to the quality of the product, a greater signaled strength of the sentiment polarity in the reviews, a lesser level of signaled uncertainty and a greater level of signaled expertise which could better explain the helpfulness of the reviews and provide more accurate predictions on most helpful reviews than the outcomes of the prior review predictions [144, 145].

$$\omega = E_f * \frac{1}{1 + e^{-\theta t_f}} \quad \text{Eq (23)}$$

They identified the nature of features of a review that might result in a high-quality or most useful review. Their input on the matter is highly informative and descriptive to our research in selecting textual qualities of reviews to gauge the quality of the reviews and in studying their effect on the performance of sentiment classification.

METHODOLOGY

The literature search and selection were done by searching high-quality research papers from journals like IEEE Access, Frontiers in Artificial Intelligence, Applied Sciences, MDPI, Springer and arXiv preprint awaiting publication. The present survey is a synthesis of the results of four high-quality and extensive surveys that have been published since 2023-2025: A total of 150 references are quoted in these four sources, which represent a comprehensive overview of the discipline, including the beginnings and the latest advancements in the field.

Data Extraction and Taxonomy Formulation

This section will extract the necessary data and formulate the taxonomy. The information that we obtained in each source was the following:

- a. Task definition and scope: Subjective language understanding.
- b. Classical ML to LLM model architectures and algorithms.
- c. Big datasets and benchmarks containing size, domain and task information.
- d. Rationalization-based explainability techniques.
- e. Measurements and strategies of prediction and explanation quality.
- f. Problems and restrictions in tasks and approaches are reported.
- g. Future research directions as identified by the domain experts.

Quality Evaluation and Integration

Each of the four source surveys has gone through peer review and publication in a reputable source (Applied Sciences, IEEE Access, Frontiers in Artificial Intelligence, arXiv preprint awaiting publication). We ensured that our findings were consistent with others conducted in the surveys, and where different authors concluded the same way and different ways. Tracing all claims to cited sources, Comprehensiveness: The coverage of tasks, models, and methods of evaluation, Recency: Including latest LLM developments through 2025, Practical utility: Emphasizing datasets, benchmarks, and evaluation methods.

The approach proposed is concerned with the application of the enhanced feature-based sentiment analysis (FbSA) to classify the reviews according to their features and provide the correct sentiment score to each feature of the reviews. The key input in the proposed method is to develop an improved feature-based sentiment analysis, which can primarily identify the spam reviews using a mixture of various properties obtained on the review itself. These are content properties, POS properties, sentiment properties and improved bag of words properties. Not just that, the method can also provide the correct sentiment scores on the features, owing to the ability to deal with negation and identify emoticons. The suggested solution consists of five major modules, i.e., (a) preprocessing, (b) features and sentiment terms extraction, (c) features sentiment scores, (d) spam reviews and (e) polarity. The overview of the proposed approach is in Figure 2.

Proposed Approach

$$\ln f_{it}^+ = \sum_{j=0}^t \Delta \ln W^T x + b_{it}^+ = \sum_{j=0}^t \max(\Delta W^T_{ij,0}) + \epsilon_{it} \quad \text{Eq (24)}$$

$$\ln f_{it}^+ = \sum_{j=1}^t \Delta \ln W^T x + b_{it}^+ = \sum_{j=1}^t \max(\Delta W^T_{ij,1}) + \epsilon_{it} \quad \text{Eq (25)}$$

The proposed solution functions in the following way: First, every review of the dataset is subjected to preprocessing. These preprocessed reviews are then transformed into features, opinion words and emoticons. Then, the polarity of each of the obtained opinion words is determined and processed by negation handling process.

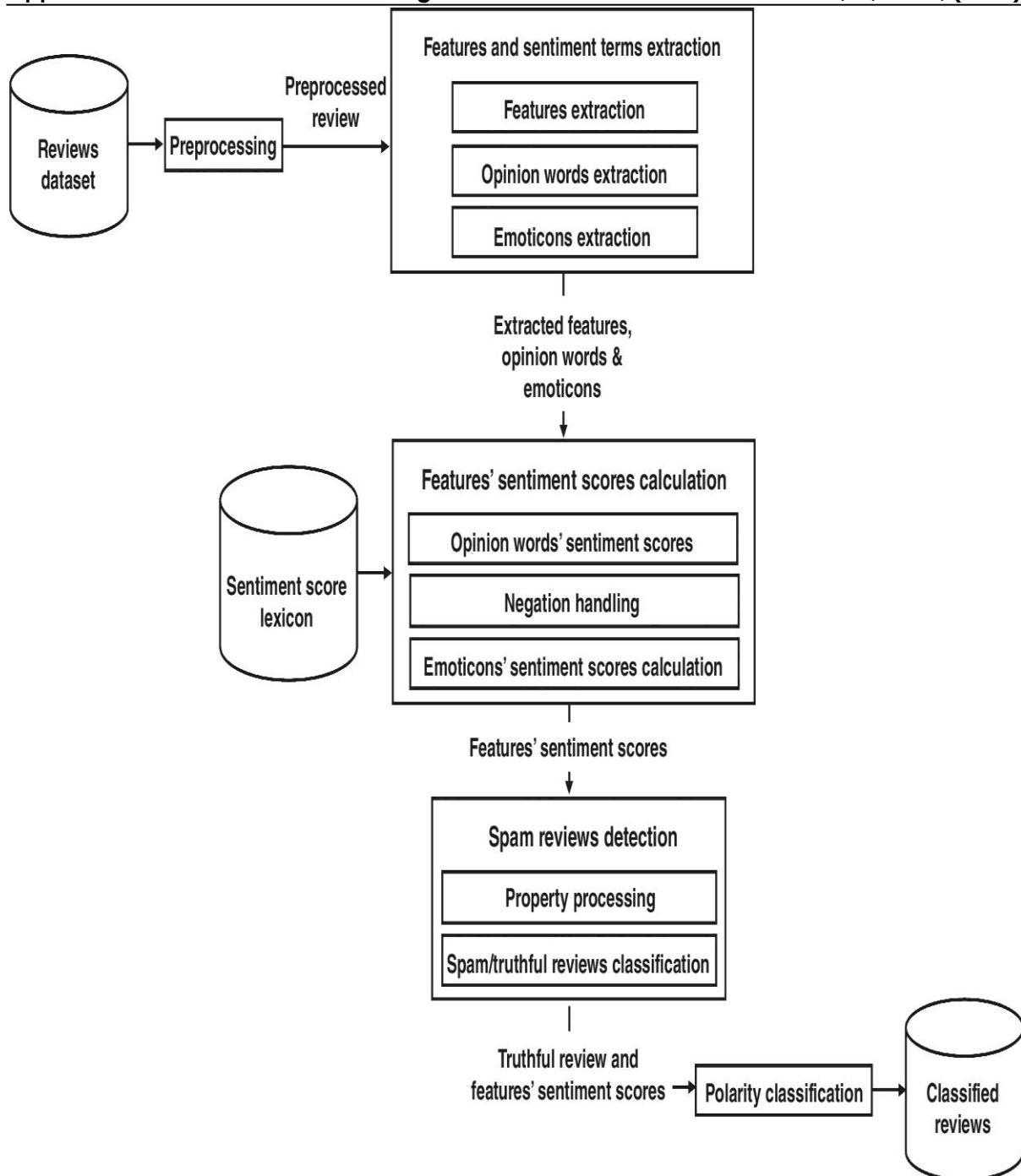


Figure 3.

The polarity of each of the extracted emoticons is then determined and subsequently sentiment score of each of the extracted features is obtained. This is then followed by a spam review detection strategy of filtering spam reviews. Lastly, every honest review is categorized as either negative or positive. In determining spam reviews, the property processing sub-module is used to find the properties that must be extracted from each review. The accuracy of the spam reviews detection module depends on the choice of the right set of properties. Various kinds of properties might be identified within the text of the reviews. This sub-module proposes various kinds of properties, which greatly influence the performance of the spam reviews detection module, to be identified in the text of the reviews.

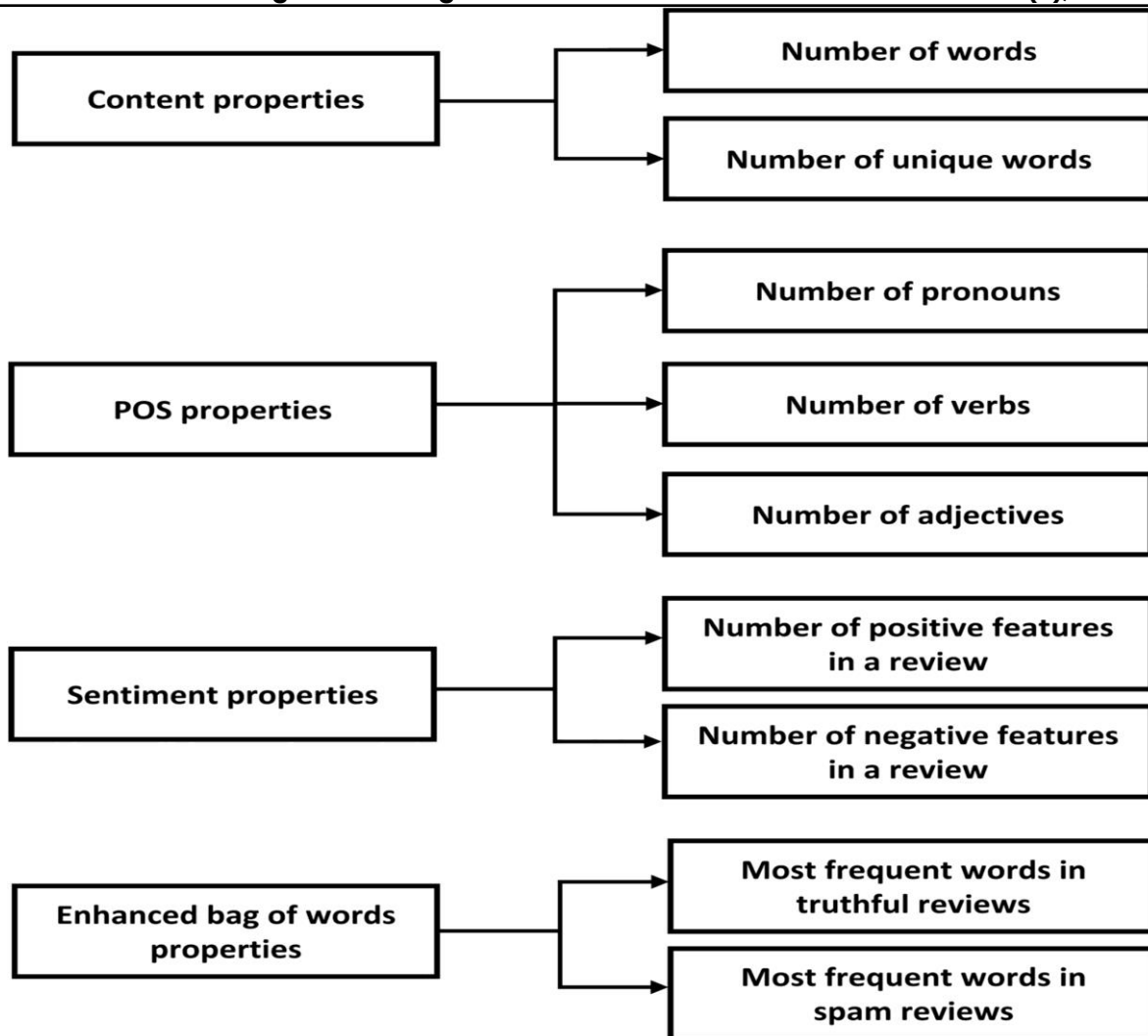


Figure 4.
Approach used for Review detection

The recommended properties are: sentiment properties and enhanced bag of words properties. This is on top of the most popular group of properties: content properties and POS properties. The content properties are properties that are based on the text of the review and the properties are divided into two sub-properties, namely number of words and number of unique words. The POS properties refer to a bunch of properties that hinge upon the frequency of POS tags within the text of the review and are made up of one subproperty, namely, number of pronouns, which is further divided into two additional subproperties, namely, number of verbs and number of adjectives. Sentiment properties. An assortment of properties that relies on the previously determined sentiment scores of features, upon neglecting negation management and emoticon identification. These properties are further subdivided into two sub-properties, namely the number of positive features in a review and the number of negative features in a review. Negation handling and emoticons detection influence these sentiment properties in a positive manner. These properties, based on this, are thus seen as being part of the key contribution in this proposed method since they greatly influence the classification of the review as either spam or truthful. Enhanced bag of words properties a collection of properties which are anchored on the common words in truthful reviews text and the common words in spam reviews text and the frequency of the common words in both texts in each review.

TAXONOMY OF SUBJECTIVE NLP TASKS

We present a unified taxonomy of eight core subjective tasks, each characterized by its definition, key challenges, and representative datasets.

Sentiment Analysis

Definition: The process of identifying and categorizing opinions or attitudes expressed in text toward an entity, product, or event. Sentiment is typically represented as polarity (positive, negative, neutral), though modern research expands to intensity scoring and aspect-based analysis. Context-dependent polarity (-not bad) is positive despite negation), Domain adaptation (movie reviews \neq , product reviews \neq tweets), Aspect-level granularity (two or more aspects with<|human|>adaptation (movie reviews \neq , product reviews \neq tweets), Aspect-level granularity (two more aspects with heterogeneous feelings), Detection of sarcasm and irony.

Modeling Evolution

Classical ML Lexicons and NB/SVM/LR achieve 80-85% accuracy on balanced datasets, Deep Learning, CNN/LSTM/BiLSTM with 85-92% accuracy, Transformers: BERT/RobERTa achieves 92-96% accuracy, LLMs, GPT-3/Llama with 85-95% zero-shot, and 90-97% with few-shot. Classical: Emotion lexicons (NRC, EmoLex) and ML classifiers, Deep Learning: BiLSTM with attention to context, Transformers: BERT fine-tuned on emotion corpora, LLMs: GPT-3/4 with emotion-labeled prompts.

Theoretical Frameworks and Sarcasm Detection

Ekman's basic emotions, joy, sadness, anger, fear, surprise, disgust, Dimensional models: valence-arousal-dominance space, Appraisal theory: cognitive evaluation of events. While it is important to determine whether an utterance is ironic or contradictory, the term here is used to refer to an utterance in which the meaning is both ironic and contradictory, literal interpretation is not the intended message. Reliance on context (must have background knowledge), An absence of overt cues in text-only data, Cultural and individual difference, Dataset quality (hashtag-based labels are noisy)

MAJOR DATASETS

Detection Methods

Semantic violation of selection preference, conceptual metaphor theory, Word embeddings represent abstract-concrete mappings, Neural: BiLSTM using contextualized representation and Transformers: BERT learns metaphorical patterns of use. Understanding the goal/purpose of communication which was behind the utterance of a user. Examples: Book me a flight to Toronto is an expressive of an intent to book a flight. Fine-grained intent taxonomies (100 more intents in banking/retail), Out-of-scope. utterance handling, Domain adaptation, and Multilingual intent classification.

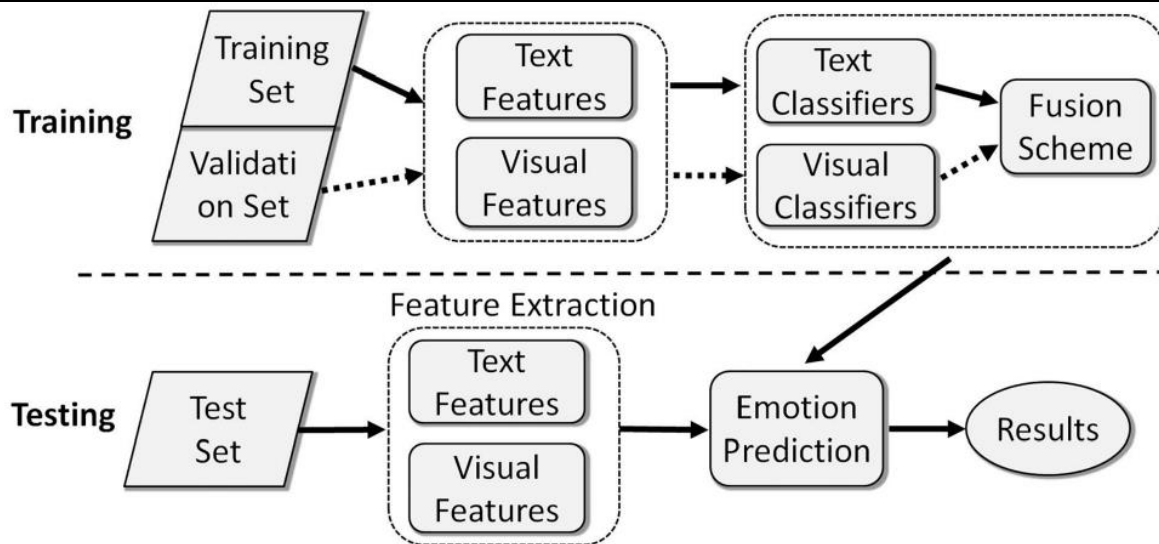


Figure 5. Dataset Training and Testing Approach
 Table 6. Major Datasets for Intent Detection

Dataset	Size	# Intents	Domain	Reference
ATIS	~5,900	26	Airline booking	[147]
SNIPS	~14,000	7	Voice assistants	[148]
BANKING77	~13,000	77	Banking	[149]
CLINC150	~23,000	150	Multi-domain	[150]

Aesthetics Assessment and Evaluation Dimensions

Definition: Evaluating the perceived quality, creativity, or beauty of text, whether of literary style, poetic appeal, or narrative coherence. High subjectivity (there is no ground truth), Multi-dimensional quality (creativity, coherence, imagery), small-scale annotated data, and Cultural and individual inclinations. Linguistic inventiveness and innovation, Coherence and narrative structure, Imagery and figurative language, Emotional impact.

Table 7. Major Datasets for Aesthetics Assessment

Dataset	Size	Aspect	Reference
Poetry Foundation	~1,000	Creativity/imagery	[151]
Writing Prompts	~300,000	Story quality	[152]
ROCStories	50,000	Narrative coherence	[152]
Story Cloze Test	3,744	Coherence	[153]

ALGORITHMS AND MODELS

In this survey, we trace the history of paradigms of subjective NLP models and discuss their implications in regard to explainability.

Traditional Machine Learning Classifiers:

Tokenization & normalization includes Lowercasing, punctuation and stop-word filtering, Lexical: TF-IDF, n-grams (unigrams, bigrams, trigrams), Sentiment: MPQA, SentiWordNet, AFINN, NRC Emotion Lexicon, Linguistic: POS tags, dependency relations, negation processing Naive Bayes Probabilistic, fast, interpretable, feature independence, Support Vector Machines Kernel methods for non-linear boundaries; strong baselines, Logistic Regression [154] Linear model with probability outputs;

interpretable coefficients, Performance: Accuracy on sentiment analysis benchmarks 75-85%, Explainability: Have a higher transparency by feature weights and lexicon scores [155]. Manual feature engineering work, Low generalization between domains, does not process figurative language and context well.

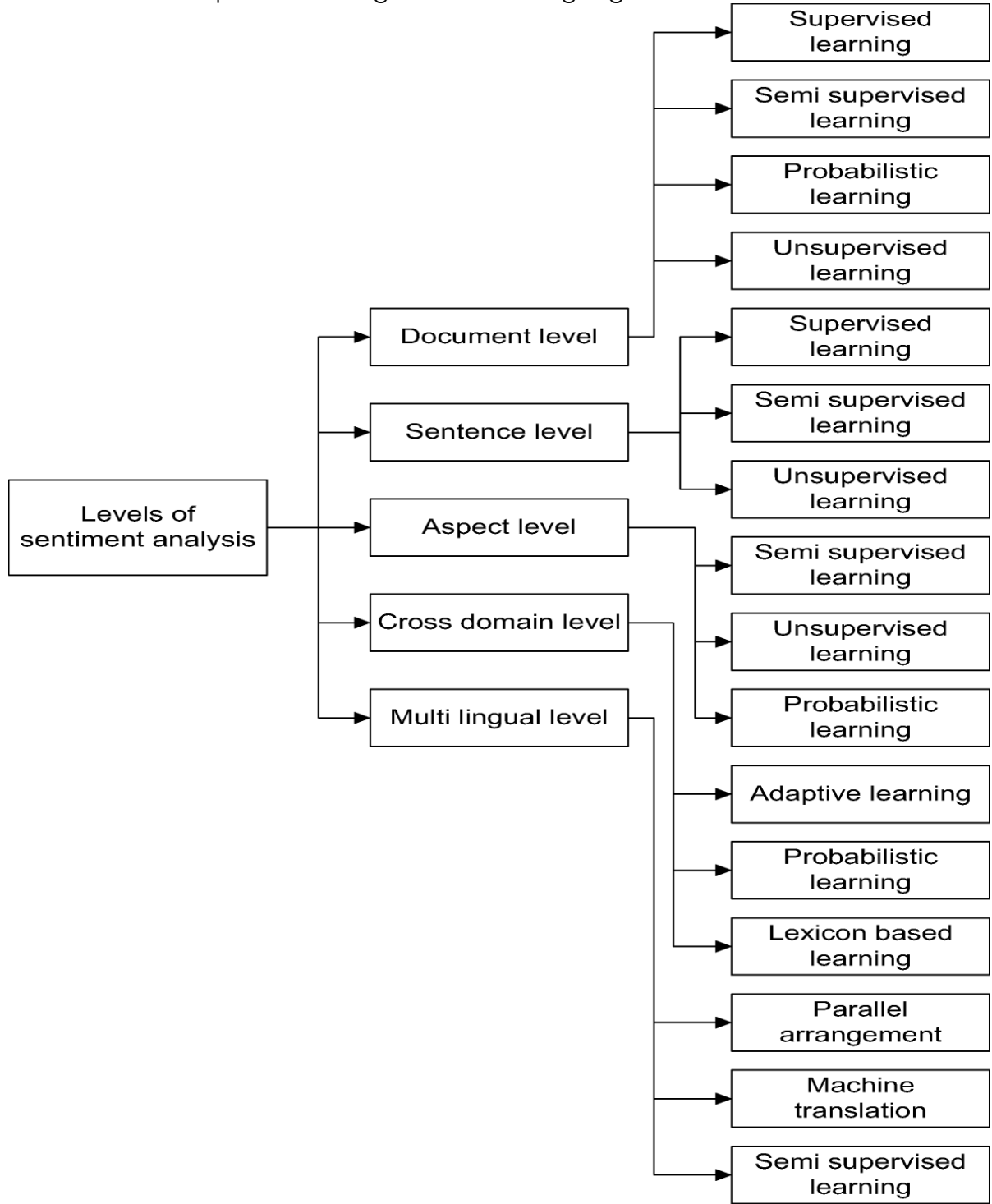


Figure 6.
Different ML approaches for sentiment analysis

The three additional sub classifications of machine learning included supervised, unsupervised and semi-supervised methods as the methods in which various learning classifiers were employed to train the dataset e.g., Decision Tree, Support Vector Machine, Neural Networks, Naive Bayes and Maximum Entropy. These were the

classifiers that were applied under the shadow of supervised approaches. Probabilistic methods and rough sets were applied on the unsupervised methods and co-training or wrapper on the semi-supervised methods. Lexicon methods further separated into two of which are dictionary and corpus based methods. Dictionary is a collection of words where each word retain dimensions such as positive, negative or can be neutral or can be created by own purpose/ event/ product/ service/ software/ field or can be made use of baseline dictionaries. Similar to the dictionary-based approach, the corpus-based approach applied to a large or big quantity of textual information, because of this factor, this method maintains a keep rate of an utterance of sentiments with possibility in the form of combination. Authors attempt to bring out the different approaches in Fig. 11 using the assistance of many researchers. These researchers pose their sentiment to be in five distinct key levels in an analysis form of document, sentence, aspect, cross domain and multi lingual sentiment.

As a result, the sentiment analysis levels were further split into sub level methods that are applied to extract the sentiments of the various kinds of data. There were five broad sentimental analysis levels which were further subclassified as supervised, unsupervised, semi-supervised, probabilistic, lexicon and adaptive based learning, parallel arrangement and machine translation. In other cases, sentiment categorization studies the dimension of a sentence the same as the dimensions of words and the same is also known as dimension examination. This mode of analysis has been running on reviews, forums, blogs, news articles and micro blogs due to the limitation in 140 words in micro blogs; it cannot inhabit sentences, therefore the limitation in the number of words lets it handle merely playing around with abbreviations amidst the cacophonous information. Data of micro blogs applied in many applications and the dimensions were used with sentiment classification.

DEEP NEURAL MODELS

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs):

Filters capture local n-gram patterns, Dimensionality reduction via max-pooling, Can be used to classify sentences at the sentence level, and CNN (2014) by Kim: Simple and effective baseline [156]. LSTM Long Short-Term Memory, GRU Gated Recurrent Units, less complex than LSTM, BiLSTM Bidirectional processing of the whole context [157].

Hybrid Models

CNN-LSTM: Local patterns and sequential modeling, Attention mechanisms: State-of-the-art pre-transformer, BiLSTM-Attention: State-of-the-art pre-transformer. Word2Vec CBOV and Skip-gram distributional semantics, GloVe: World vectors based on co-occurrence statistics, FastText: Sub-word embeddings based on morphology. Ranges between 85 to 92% for sentiment analysis; which is massive improvement over classical ML. Smaller than classical ML; attention weights give partial interpretability. Sequential processing (slow on long texts), Difficulty capturing long-range dependencies, and Less interpretable than feature-based methods

Self-Attention Revolution Sequence parallel processing, Multi-head attention to represent many different features, Positional encodings of the order of a sequence, and Scaled dot-product attention mechanism.

Table 8.
BERT Family with its Major Innovation and Parameters

Approach based on Machine Learning**Ahmed, A, et al., (2025)**

Model	Parameters	Key Innovation	Reference
BERT-base	110M	Bidirectional MLM + NSP	[157]
BERT-large	340M	Larger capacity	[158]
RoBERTa	355M	Optimized training	[159]
DistilBERT	66M	Knowledge distillation	[160]
ALBERT	12M-235M	Parameter sharing	[161]
ELECTRA	110M	Discriminative pretraining	[162]
DeBERTa	1.5B	Disentangled attention	[163]

Pretraining: Pretraining Masked Language Modeling (MLM) on large corpora, and Fine-tuning: Heads on fine-tuning classification. T5 (Text-to-Text Transfer Transformer): Universal text-to-text structure, and BART: Denoising autoencoder generation. 90-96 percent of sentiment analysis; new state-of-the-art among subjective tasks. Attention visualization, Probing classifiers for linguistic knowledge, Layer-wise analysis of representations, still requires post-hoc interpretation methods [164].

Large Language Models: Zero/Few-Shot Subjectivity

GPT Family (Decoder-only)

Table 9.**Open Ai's GPT version along with its training data and parameters**

Model	Parameters	Training Data	Context	Reference
GPT	117M	BooksCorpus	512	[164]
GPT-2	1.5B	WebText	1,024	[165]
GPT-3	175B	Common Crawl + curated	2,048	[166]
GPT-4	Unknown (>1T est.)	Multimodal	32k/128k	[167]

Open-Source LLMs Llama (7B-65B): Efficient open foundation models, Llama 2 (7B-70B): Chat tuned models are improved, PaLM (540B): Pathways Language Model, and Falcon, MPT, Mistral: Community alternatives. Explaining reasoning

Table 10.**Major Sentiment Analysis Benchmarks [168]**

Dataset	Size	Domain	Classes	Language	Annotation
IMDb	50,000	Movie reviews	Binary	English	User ratings
Sentiment140	1.6M	Twitter	Binary	English	Emoticon-based
Twitter	US 14,640	Customer service	3-class	English	Crowdsourced
Airline					
SemEval-2017 Task 4	~50,000	Twitter	Multi-subtask	English/Arabic	Expert
Amazon Reviews	142M	E-commerce	5-star	English	User ratings
Yelp Reviews	1.6M+	Business/restaurant	5-star	English	User ratings
SST (Stanford)	11,855	Movie reviews	5-class/binary	English	Expert

Table 11.**Emotion Recognition Benchmarks [169]**

Dataset	Size	# Emotions	Domain	Modality	Classes
ISEAR	7,666	7 basic	Survey responses	Text	Binary
GoEmotions	58,000	27 fine-grained	Reddit comments	Text	Binary
SemEval-2018 Task 1	~10,000	11 emotions	Twitter	Text	3-class
DailyDialog	13,118	7 basic	Conversations	Text	Multi-subtask
MELD	13,000	7 basic	TV dialogues	Multimodal	5-star
EmoBank	10,000	VAD dimensions	Various	Text	5-star

Emotion Taxonomies

Basic emotions: joy, sadness, anger, fear, surprise, disgust, neutral
 Fine-grained: admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise, neutral, Valence (positive-negative), Arousal (activated-deactivated), Dominance (controlled-in-control)

Table 12. Extractive Rationalization Evaluation Metrics

Metric	Definition	Measures
Token F1	Overlap with human rationales	Plausibility
IOU (Intersection over Union)	Span overlap measure	Plausibility
Comprehensiveness	Prediction change when removing rationale	Fidelity (sufficiency)
Sufficiency	Prediction maintained with rationale only	Fidelity
Aopc (Area over Perturbation Curve)	Cumulative prediction degradation	Fidelity

Table 13. Abstractive Rationalization Evaluation

Metric	Definition	Measures
BLEU/ROUGE	N-gram overlap with gold explanations	Surface similarity
BERTScore	Semantic similarity using embeddings	Semantic quality
Human rating	Expert judgment of quality	Plausibility
Faithfulness score	Explanation aligns with model internals	Fidelity
Simulation consistency	Humans predict label from explanation	Comprehensibility

Table 14. Rationalization Benchmark Datasets

Dataset	Task	Size	Rationale Type
e-SNLI	NLI	570k	Abstractive
ERASER	Multi-task	~175k	Extractive
CoS-E	Commonsense QA	10k	Abstractive
FEVER	Fact verification	185k	Extractive (evidence)
MultiRC	Reading comp.	10k	Extractive
Movie Reviews	Sentiment	2k	Extractive
BeerAdvocate	Aspect sentiment	1.5k	Extractive

Challenges in Rationalization

No consensus on best evaluation protocol, and Fidelity and plausibility sometimes conflict, Human evaluation expensive and variable. Models may rely on dataset artifacts, Rationales highlight spurious features (e.g., negation words for positive class), and Adversarial robustness of explanations. Attribution methods require many forward passes, LLM explanation generation adds latency, Trade-off between explanation quality and efficiency. Audio-visual-text integration for sarcasm/humor, How to generate unified multimodal explanations?, Limited benchmarks and methods

COMPARATIVE ANALYSIS AND INTEGRATION

Cross-Task Patterns and Explainability Trade-offs Across Model Families

Analyzing performance across the eight subjective tasks reveals consistent patterns

Table 15.

Model Performance Summary Across Tasks

Task	Classical ML	Deep Learning	Transformers	LLMs (Zero-shot)	LLMs (Few-shot)
Sentiment Analysis	75-85%	85-92%	92-96%	85-95%	90-97%
Emotion Recognition	70-80%	80-88%	88-93%	70-85%	80-90%
Sarcasm Detection	60-70%	70-80%	80-88%	60-75%	70-82%
Humor Detection	55-65%	65-75%	75-85%	65-80%	72-85%
Stance Detection	65-75%	75-83%	83-90%	75-88%	82-92%
Metaphor Recognition	60-72%	72-82%	82-90%	70-85%	78-90%
Intent Detection	80-88%	88-93%	93-97%	80-92%	88-95%
Aesthetics Assessment	50-60%	60-70%	70-82%	65-78%	70-83%

Table 16.
Explainability Characteristics by Model Family

Model Family	Accuracy	Inherent Interpretability	Post-hoc Explainability	Rationalization Support
Classical ML	Low-Medium	High (feature weights)	Low need	Lexicon-based
Deep Neural	Medium	Low (hidden states)	Medium (attention)	Attention-based
Transformers	High	Low (distributed)	Medium (attribution)	Extractive modules
LLMs	High	Very Low (scale)	High (generation)	Abstractive prompting

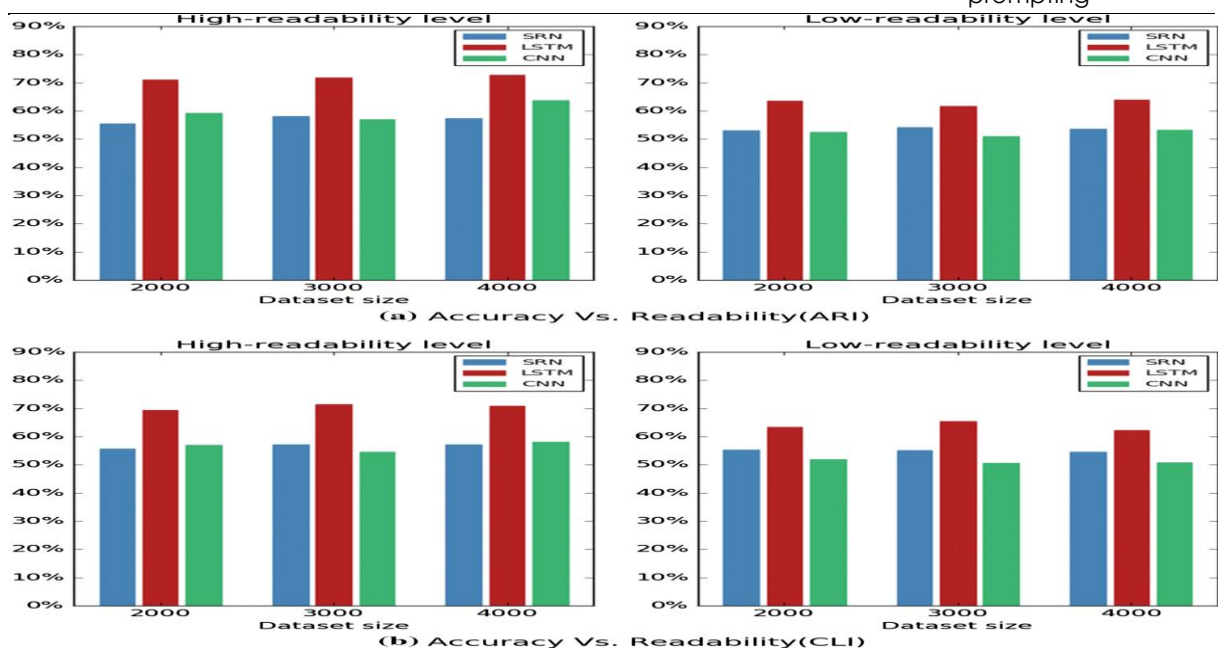


Figure 7.
Accuracy and Readability for Different Models (ARI) (b) Accuracy and Readability for Different Models (CLI)

CONCLUSION

The articles considered in this paper are categorized into six broad areas of the sentiments that include, analysis, classification, feature selection, emotion detection and transfer learning and building resources. The identified division assists the review on sentiment analysis to receive the aspects such as in how problems should be addressed, datasets and methods/structures/techniques use, feature selection, approaches, empirical tests, results and explore the prospect enrichment. The datasets, properties and preprocessing steps of the same are addressed in various articles. These datasets have been utilized to increase the performance of empirical tests. The degrees of sentiments namely document, sentence, aspect, cross domain

and multi lingual etc are pointed out. The publicly available datasets, as well as toolkits that are helpful in various lexicon and corpus based methods, are also detailed on the other hand, this study is experiencing some significant community medium controls and test examination parameters that are generally utilized under examination and classification with good influence. Typically, there are four major categories of approaches that include supervised, unsupervised, semi supervised and hybrid. The approaches, levels, datasets and toolkits available, the properties, the controls of the sources of communication and mediums of the community and test evaluation parameters make the major part of this systematic research. Thus, in the case of sentiment classification area, newcomers will find this paper handy. In the end, the way to go is self-explanatory NLP systems that will bring together subjective and these systems should:

1. Accurately predict on various subjective tasks
2. Explain transparently using natural language rationales
3. Verify faithfully through causal intervention tests
4. Serve equitably across languages, cultures, and demographics
5. Deploy sustainably with manageable computational costs

Integration of LLM abilities, rationalization and strict assessment standards sets the field to achieve such a vision. Future research should give more importance to standardized benchmarks, causal explainability, multimodal reasoning, and human. Feedback, multiculturalism, computational performance, and ethics governance. By integrating subjective language understanding with explainability, we enable machines not only to predict what humans feel or intend but also to explain why. Transparently, responsibly, and across cultures. This single framework drives forward science. Knowledge and skills in the practical implementation of NLP systems, which can be used by human beings to serve their needs. Creditworthiness and responsibility. The contemporary healthcare system produces enormous amounts of data from various sources. They are electronic health records (EHRs), medical imaging systems, lab results, wearable health monitors, mobile health apps, insurance claims and, most recently, genomics data.

It is this heterogeneous combination of structured and unstructured data, which encompasses the patient demographics and diagnostic codes, but goes all the way up to the continuous biometric streams to what is known as big data in healthcare. Big data presents radical possibilities in medicine. It allows making diagnoses more accurately, supporting predictive analytics, providing patients with an opportunity to detect diseases at an early stage, managing the health of the population, and improving clinical decision-making. Nevertheless, storing and controlling such large volumes of sensitive data is extremely demanding in terms of technical and ethical factors. The healthcare system's big data lifecycle can be generally broken down into five steps that are essential: data collection, storage, analysis, utilization, and destruction. There are vulnerabilities presented by each stage. As an example, when collecting data, the data sent by IoT-based devices, such as fitness trackers or smart medical sensors, may be compromised because of the insecure transmission protocols. As soon as they have been gathered, the information is stored in centralized or cloud-based repositories that are frequently targeted by cybercriminals because of the high price of medical records.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the contributor to the research and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.

Availability of data and material: In the approach, the data sources for the variables are stated.

Authors' contributions: Each author participated equally in the creation of this work.

Conflicts of Interest: The authors declare no conflict of interest.

Consent to Participate: Yes

Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- Abdullah, M. M., Ghafoor, U., Qadeer, Q. B., Khadim, F., Khan, H. S., Ahmad, A., & Khan, H. (2025). An Efficient of Artificial Intelligence based Brain Tumor Diagnosis and Classification: An Advance Medical Diagnosis Approach. *The Asian Bulletin of Big Data Management*, 5(2), 208-242.
- Abdullah, M. M., Ghafoor, U., Qadeer, Q. B., Khadim, F., Khan, H. S., Ahmad, A., & Khan, H. (2025). An Efficient of Artificial Intelligence based Brain Tumor Diagnosis and Classification: An Advance Medical Diagnosis Approach. *The Asian Bulletin of Big Data Management*, 5(2), 208-242.
- Abdullah, M. M., Khan, H., Farhan, M., & Khadim, F. (2024). An Advance Machine Learning (ML) Approaches for Anomaly Detection based on Network Traffic. *Spectrum of engineering sciences*, 2(3), 502-527.
- Abouelmehdi, K., Beni-Hssane, A., Khaloufi, H., & Saadi, M. (2017). Big data security and privacy in healthcare: A Review. *Procedia Computer Science*, 113, 73-80.
- Adadi, Amina, and Mohammed Berrada. "Explainable AI for Healthcare." *Artificial Intelligence Review*, Springer, 2020.
- Adil, M. U., Ali, S., Haider, A., Javed, M. A., & Khan, H. (2024). An Enhanced Analysis of Social Engineering in Cyber Security Research Challenges, Countermeasures: A Survey. *The Asian Bulletin of Big Data Management*, 4(4), 321-331.
- Ahmad, I., Nasim, F., Khawaja, M. F., Naqvi, S. A. A., & Khan, H. (2025). Enhancing IoT Security and Services based on Generative Artificial Intelligence Techniques: A Systematic Analysis based on Emerging Threats, Challenges and future Directions. *Spectrum of engineering sciences*, 3(2), 1-25.
- Ahmad, J., Salman, W., Amin, M., Ali, Z., & Shokat, S. (2024). A Survey on Enhanced Approaches for Cyber Security Challenges Based on Deep Fake Technology in Computing Networks. *Spectrum of Engineering Sciences*, 2(4), 133-149.
- Akmal, I., Khan, H., Khushnood, A., Zulfiqar, F., & Shahbaz, E. (2024). An Efficient Artificial Intelligence (AI) and Blockchain-Based Security Strategies for Enhancing the Protection of Low-Power IoT Devices in 5G Networks. *Spectrum of engineering sciences*, 2(3), 528-586.
- Al Zaabi, M., & Alhashmi, S. M. (2024). Big data security and privacy in healthcare: A systematic review and future research directions. *Information Development*, 02666669241247781.
- Ali, A. (2019). Intelligent Auto Traffic Signal Controller for Emergency Vehicle by Using. *Journal of Engineering and Applied Sciences*, 14(1), 76-82.
- Ali, Arshad. "Images classification and feature extraction by using unmanned aerial vehicles (uav's)." *International Journal of Applied Engineering Research* 12.24 (2017): 15432-15437.

- Ali, G., Shahbaz, H., Hassan, M. A., Ahmad, M., & Waleed, M. (2024). An Enhanced Approach of Exploring Digital Economy Using Modern Computer Networks. *Spectrum of Engineering Sciences*, 2(4), 292-312.
- Ali, H., Ayub, N., Irfan, A., Fayyaz, S., Masood, H., Ahmad, A., ... & Khan, H. (2025). A Unified AI-powered Social Media Platform for Intelligent Scheduling and Data Driven Analytics Using Multi-Layered Artificial Neural Networks (ANNs): <https://doi.org/10.5281/zenodo.17572988>. *Annual Methodological Archive Research Review*, 3(11), 94-134.
- Ali, I., Saleem, M. U., Khan, A. A., Naz, A., Nawaz, M., & Khan, H. (2025). An Enhanced Artificial Intelligence Generated Virtual Influencer Framework: Examining the Effects of Emotional Display on User Engagement based on Convolutional Neural Networks (CNNs). *Annual Methodological Archive Research Review*, 3(4), 184-209.
- Almalawi, A., Khan, A. I., Alsolami, F., Abushark, Y. B., & Alfakeeh, A. S. (2023). Managing security of healthcare data for a modern healthcare system. *Sensors*, 23(7), 3612.
- Alsubaei, Faisal, et al. "Security and Privacy in IoT Healthcare: Review." *Applied Sciences*, MDPI, 2021.
- Alzu'bi, Ahmad, et al. "Privacy and Edge Computing in Smart Healthcare." *ResearchGate*, 2024.
- Anas, M., Imtiaz, M. A., Saad Khan, A. A., Naghman, N. F., Khan, H., & Albouq, S. AN ADVANCED MACHINE LEARNING (ML) ARCHITECTURE FOR HEART DISEASE DETECTION, PREDICTION AND CLASSIFICATION USING MACHINE LEARNING. Vol.-20, No.-3, March (2025) pp 54 - 72
- Aqeel, I., et al. "IoT Smart Medical Devices in Healthcare." *IEEE Access*, 2021.
- Aqeel, N., Alam, A., Bhatti, Z., & Amir, A. (2024). A Survey on Tor's Multi Layer Architecture and Web Implications in Dark Web. *Spectrum of Engineering Sciences*, 2(4), 212-231.
- Asghar, M. A., Aslam, A., Bakhet, S., Saleem, M. U., Ahmad, M., Gohar, A., & Khan, H. (2025). An Efficient Integration of Artificial Intelligence-based Mobile Robots in Critical Frames for the Internet of Medical Things (IoMTs) Using (ADP2S) and Convolutional Neural Networks (CNNs). *Annual Methodological Archive Research Review*, 3(4), 160-183.
- Aslam, I., Tariq, W., Nasim, F., Khan, H., Khawaja, M. F., Ahmad, A., & Nawaz, M. S. (2025). A Robust Hybrid Machine Learning based Implications and Preventions of Social Media Blackmailing and Cyber bullying: A Systematic Approach.
- Ayub, N., Alghamdi, T., Din, I., Ali, A., Khan, H., Ganiyeva, O., & Makhmudov, S. (2025). An Enhanced Artificial Intelligence and Deep Learning Assisted Breast Cancer Classification and Diagnosis Based on the Internet of Medical Things (IoMTs). *Engineering, Technology & Applied Science Research*, 15(6), 30612-30616.
- Ayub, N., Alghamdi, T., Din, I., Ali, A., Khan, H., Ganiyeva, O., & Makhmudov, S. (2025). An Enhanced Artificial Intelligence and Deep Learning Assisted Breast Cancer Classification and Diagnosis Based on the Internet of Medical Things (IoMTs). *Engineering, Technology & Applied Science Research*, 15(6), 30612-30616.
- Aziz, R., Mehmood, A., Tariq, A., Nasim, F., Farooq, U., Naqvi, S. A. A., & Khan, H. (2025). Critical Evaluation of Data Privacy and Security Threats: An Intelligent Federated Learning-based Intrusion Detection System Poisoning Attack and Defense for Cyber-Physical Systems its Issues and Challenges Related to Privacy and Security in IoT. *The Asian Bulletin of Big Data Management*, 5(1), 73-84.
- Bacha, A., Sehar, H., Naseem, S., & Khan, M. I. (2024). FEDERATED LEARNING FOR THREAT INTELLIGENCE SHARING: A PRIVACY-PRESERVING COLLABORATIVE DEFENSE MODEL. *Spectrum of Engineering Sciences*, 656-664.
- Badr, Yasmine. "The Use of Big Data in Personalized Healthcare." *Frontiers in Medicine*, 2024.
- Criado, M.F.; Casado, F.E.; Iglesias, R.; Regueiro, C.V.; Barro, S. Non-iid data and continual learning processes in federated learning: A long road ahead. *Inf. Fusion* 2022, 88, 263–280.
- Fakhar, M. H., Baig, M. Z., Ali, A., Rana, M. T. A., Khan, H., Afzal, W., ... & Albouq, S. (2024). A Deep Learning-based Architecture for Diabetes Detection, Prediction, and Classification. *Engineering, Technology & Applied Science Research*, 14(5), 17501-17506.
- Farooq, I., Ahmed, S. A., Ali, A., Warraich, M. A., Aqeel, M., & Khan, H. (2024). Enhanced

- Classification of Networks Encrypted Traffic: A Conceptual Analysis of Security Assessments, Implementation, Trends and Future Directions. *The Asian Bulletin of Big Data Management*, 4(4), 500-522.
- Farooq, M., Younas, R. M. F., Qureshi, J. N., Haider, A., & Nasim, F. (2025). Cyber security risks in DBMS: Strategies to mitigate data security threats: A systematic review. *Spectrum of engineering sciences*, 3(1), 268-290.
- Fatima, M., Ali, A., Ahmad, M., Nisa, F. U., Khan, H., & Raheem, M. A. U. Enhancing The Resilience Of Iot Networks: Strategies And Measures For Mitigating Ddos Attacks. *Cont.& Math. Sci.*, Vol.-19, No.-10, 129-152, October 2024 <https://jmcms.s3.amazonaws.com/wp-content/uploads/2024/10/10072102/jmcms-2410025-ENHANCING-THE-RESILIENCE-OF-IOT-NETWORKS-MF-HK.pdf>
- Fawy, K. F., Rodriguez-Ortiz, G., Ali, A., Jadeja, Y., Khan, H., Pathak, P. K., ... & Rahman, J. U. (2025). Catalytic exploration metallic and nonmetallic nano-catalysts, properties, role in photoelectrochemistry for sustainable applications. *Reviews in Inorganic Chemistry*, (0).
- Ghafoor, U., Ayub, N., Yaseen, A., Anas, M., Farooq, I., Khan, S., & Naghman, N. F. (2025). AI Assisted Heart Disease Prediction and Classification and Segmentation based on PIMA and UCI Machine Learning Datasets. *Annual Methodological Archive Research Review*, 3(7), 248-276.
- Gul, W., Nawaz, A., Hamaz, M. T., & Khan, H. AN EFFICIENT MODEL FOR THE SELECTION OF LEADERSHIP COMPETENCIES AND PERFORMANCE IMPROVEMENT FOR THE SUCCESS OF TRANSPORTATION PROJECTS, *JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES* Vol.-16, No.-5, May (2021) pp 49-65 <https://doi.org/10.26782/jmcms.2021.05.00005>
- Gularte, K.H.M.; Vargas, J.A.R.; Da Costa, J.P.J.; Da Silva, A.A.S.; Santos, G.A.; Wang, Y.; Müller, C.A.; Lipps, C.; Júnior, R.T.S.; Vidal Filho, W.B.; et al. Safeguarding the V2X Pathways: Exploring the Cybersecurity Landscape through Systematic Literature Review. *IEEE Access* 2024, 12, 72871–72895.
- Gupta, H., et al. (2021). 'Secure Cloud Storage for Healthcare using Blockchain.' *Journal of Medical Systems*.
- H. Khan, I. Uddin, A. Ali, M. Husain, "An Optimal DPM Based Energy-Aware Task Scheduling for Performance Enhancement in Embedded MPSoC", *Computers, Materials & Continua.*, vol. 74, no. 1, pp. 2097-2113, Sep. 2023
- H. Khan, M. U. Hashmi, Z. Khan, R. Ahmad, "Offline Earliest Deadline first Scheduling based Technique for Optimization of Energy using STORM in Homogeneous Multi-core Systems", *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 18, no. 12, pp. 125-130, Dec. 2018
- Hamayun Khan, Sheeraz Ahmed, S. Farhan Haider Shah, Rehan Ali Khan, Zeeshan Najam, Hasnain Abbas, Asif Nawaz, Zubair Aslam Khan, *JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES* Vol.-15, No.-8, August (2020) pp 628-646 <https://doi.org/10.26782/jmcms.2020.08.00053>
- Hasan, Md. K., et al. "Big Data and IoT: Security and Storage Challenges." arXiv, 2021.
- Hashmi, U., & Zeeshan Najam, S. A. (2023). Thermal-Aware Real-Time Task Schedulability test for Energy and Power System Optimization using Homogeneous Cache Hierarchy of Multi-core Systems. *Journal of Mechanics of Continua and Mathematical Sciences*, 14(4), 442-452.
- Hassan, A., Khan, H., Ali, A., Sajid, A., Husain, M., Ali, M., ... & Fakhar, H. (2024). An Enhanced Lung Cancer Identification and Classification Based on Advanced Deep Learning and Convolutional Neural Network. *Bulletin of Business and Economics (BBE)*, 13(2), 136-141.
- Hassan, H. Khan, I. Uddin, A. Sajid, "Optimal Emerging trends of Deep Learning Technique for Detection based on Convolutional Neural Network", *Bulletin of Business and Economics (BBE)*, vol. 12, no. 4, pp. 264-273, Nov. 2023
- He, Y., Yu, F. R., Zhao, N., Yin, H., Yao, H., & Qiu, R. C. (2016). Big data analytics in mobile cellular networks. *IEEE access*, 4, 1985-1996.
- HIPAA (1996). Health Insurance Portability and Accountability Act. U.S. HHS.
- Hornyack, P., et al. (2011). 'AppFence: Toward usable privacy for Android.' *USENIX Security Symposium*.

- Hossain, M., et al. "IoT-Enabled Teleconsultation in Healthcare." Springer Telemedicine, 2021.
- Hu, J., & Vasilakos, A. V. (2016). Energy big data analytics and security: challenges and opportunities. *IEEE Transactions on Smart Grid*, 7(5), 2423-2436.
- Hussain, M., Ahmed, H. A., Babar, M. Z., Ali, A., Shahzad, H. M., Rehman, S. U., ... & Alshahrani, A. M. (2025). An Enhanced Convolutional Neural Network (CNN) based P-EDR Mechanism for Diagnosis of Diabetic Retinopathy (DR) using Machine Learning. *Engineering, Technology and Applied Science Research*, 15(1), 19062-19067.
- Hussain, S., Sarwar, N., Ali, A., Khan, H., Din, I., Alqahtani, A. M., ... & Ali, A. (2025). An Enhanced Random Forest (ERF)-based Machine Learning Framework for Resampling, Prediction, and Classification of Mobile Applications using Textual Features. *Engineering, Technology & Applied Science Research*, 15(1), 19776-19781.
- Imtiaz, M. A., Amir, A., Bakhet, S., Siddique, H., & Rizwan, S. M. (2025). An Optimal Diabetic Retinopathy Detection and Classification Approach based on integrated Hybrid Convolutional Neural Networks (CNNs). *Spectrum of Engineering Sciences*, 3(2).
- Islam, S. M. Riazul, et al. "The Role of IoMT and Big Data in Modern Healthcare." *Journal of Network and Computer Applications*, Elsevier, 2022.
- Ismaeel, S., Saleemi, H., Amir, U., Ashraf, S., & Hamza, A. (2024). A Detailed Review of latest Trends, Technologies Applications
- Jabeen, T., Mehmood, Y., Khan, H., Nasim, M. F., & Naqvi, S. A. A. (2025). Identity Theft and Data Breaches How Stolen Data Circulates on the Dark Web: A Systematic Approach. *Spectrum of engineering sciences*, 3(1), 143-161.
- Jabeen, T., Mehmood, Y., Khan, H., Nasim, M.F. and Naqvi, S.A.A., 2025. Identity Theft and Data Breaches How Stolen Data Circulates on the Dark Web: A Systematic Approach. *Spectrum of engineering sciences*, 3(1), pp.143-161.
- Javed, M. A., Ahmad, M., Ahmed, J., Rizwan, S. M., & Tariq, A. (2025). An Enhanced Machine Learning based Data Privacy and Security Mitigation Technique: An Intelligent Federated Learning (FL) Model for Intrusion Detection and Classification System for Cyber-Physical Systems in Internet of Things (IoTs). *Spectrum of Engineering Sciences*, 3(2), 377-401.
- Javed, M. A., Anjum, M., Ahmed, H. A., Ali, A., Shahzad, H. M., Khan, H., & Alshahrani, A. M. (2024). Leveraging Convolutional Neural Network (CNN)-based Auto Encoders for Enhanced Anomaly Detection in High-Dimensional Datasets. *Engineering, Technology & Applied Science Research*, 14(6), 17894-17899.
- Khan, A. Ali, S. Alshmrany, "Energy-Efficient Scheduling Based on Task Migration Policy Using DPM for Homogeneous MPSoCs", *Computers, Materials & Continua.*, vol. 74, no. 1, pp. 965-981, Apr. 2023
- Khan, A. K., Bakhet, S., Javed, A., Rizwan, S. M., & Khan, H. (2025). Framework for Predicting Customer Sentiment Aware Queries and Results in Search Using Oracle and Machine Learning. *Spectrum of Engineering Sciences*, 3(2), 588-617.
- Khan, K. Janjua, A. Sikandar, M. W. Qazi, Z. Hameed, "An Efficient Scheduling based cloud computing technique using virtual Machine Resource Allocation for efficient resource utilization of Servers", In *2020 International Conference on Engineering and Emerging Technologies (ICEET)*, IEEE., pp. 1-7, Apr. 2020
- Khawar, M. W., Ayub, N., Shaheen, S., Iffikhar, B., Masood, H., Ahmad, A., & Khan, H. (2025). An Efficient system based on Artificial Intelligence for the Detection and Mitigation of network Intrusion using encrypted traffic protocols: A Systematic Approach. *Annual Methodological Archive Research Review*, 3(11), 32-71.
- Khawar, M. W., Salman, W., Shaheen, S., Shakil, A., Iffikhar, F., & Faisal, K. M. I. (2024). Investigating the most effective AI/ML-based strategies for predictive network maintenance to minimize downtime and enhance service reliability. *Spectrum of Engineering Sciences*, 2(4), 115-132.
- Kumar, Neeraj, et al. "IoT Security Challenges in 5G Healthcare." Springer, 2020.
- Li, H.; Luo, L.; Wang, H. Federated learning on non-independent and identically distributed data. In *Proceedings of the Third International Conference on Machine Learning and Computer Application (ICMLCA 2022)*, Shenyang, China, 16–18 December 2023; SPIE: Bellingham, WA, USA; pp. 154–162.

- Liang, Y., Ur Rahman, S., Shafaqat, A., Ali, A., Ali, M. S. E., & Khan, H. (2024). Assessing sustainable development in E-7 countries: technology innovation, and energy consumption drivers of green growth and environment. *Scientific Reports*, 14(1), 28636.
- Liaqat, M. S., Sharif, N., Ali, A., Khan, H., Ahmed, H. N., & Khan, H. (2024). An Optimal Analysis of Cloud-based Secure Web Applications: A Systematic Exploration based on Emerging Threats, Pitfalls and Countermeasures. *Spectrum of engineering sciences*, 2(5), 427-457.
- Liu, Y., et al. (2020). 'Shadow Coding for Privacy-Preserving Data Aggregation.' *IEEE Transactions*.
- M. Gondal, Z. Hameed, M. U. Shah, H. Khan, "Cavitation phenomenon and its effects in Francis turbines and amassed adeptness of hydel power plant", In 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE., pp. 1-9, Mar. 2019
- Mahmood, F., Shehroz, M., Ansari, Z., & Rauf, F. (2024). A Survey of Software-Defined Networks Based on Advance Machine Learning Based Techniques. *Spectrum of Engineering Sciences*, 2(4), 232-257.
- Maqsood, M., Dar, M. M., Javed, M. A., & Khan, H. (2024). A Survey on the Internet of Medical Things (IOMT) Privacy and Security: Challenges Solutions and Future from a New Perspective. *The Asian Bulletin of Big Data Management*, 4(4), 355-368.
- Mittal, R., et al. (2018). 'Homomorphic Encryption in Cloud Health Environments.' *ACM Computing Surveys*.
- Muhammad Anas, Muhammad Atif Imtiaz, Saad Khan, Arshad Ali, Noor Fatima Naghman, Hamayun Khan, Sami Albouq, AN ADVANCED MACHINE LEARNING (ML) ARCHITECTURE FOR HEART DISEASE DETECTION, PREDICTION AND CLASSIFICATION USING MACHINE LEARNING, *Cont. & Math. Sci*, Vol.20, No.3, 2025 <https://doi.org/10.26782/jmcms.2025.03.00005>
- Mujtaba, A., Zulfqar, M., Azhar, M. U., Ali, S., Ali, A., & Khan, H. (2025). ML-based Fileless Malware Threats Analysis for the Detection of Cyber security Attack based on Memory Forensics: A Survey. *The Asian Bulletin of Big Data Management*, 5(1), 1-14.
- Mumtaz, J., Bakhet, S., Javed, A., Naz, A., Rashail, M., & Khan, H. (2025). An Intelligent Diagnosis and Tumor Segmentation Method based on MRI Images Using Pre-trained Deep Convolutional Neural Networks (CNNs). *The Asian Bulletin of Big Data Management*, 5(1), 147-163
- Mumtaz, J., Rehman, A. U., Khan, H., Din, I. U., & Tariq, I. Security and Performance Comparison of Window and Linux: A Systematic Literature Review. *Securing the Digital Realm*, 272-280.
- Mustafa, M., Ali, M., Javed, M. A., Khan, H., Iqbal, M. W., & Ruk, S. A. (2024). Berries of Low-Cost Smart Irrigation Systems for Water Management an IoT Approach. *Bulletin of Business and Economics (BBE)*, 13(3), 508-514.
- F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.
- A. Atanasova, J. G. Simonsen, C. Lioma, and I. Augenstein, "Generating label cohesive and well-formed adversarial claims," in *Proc. EMNLP*, 2020, pp. 3168–3177.
- S. Wiegrefe and A. Marasović, "Teach me to explain: A review of datasets for explainable natural language processing," in *Proc. NeurIPS Datasets and Benchmarks Track*, 2021.
- R. Caruana, "Multitask learning," *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997.
- Y. Zhang, X. Wei, B. Y. Lin, Y. F. Wu, and Y. Xu, "Neural relation extraction via inner-sentence noise reduction and transfer learning," in *Proc. EMNLP*, 2018, pp. 2195–2204.
- B. Mittelstadt, C. Russell, and S. Wachter, "Explaining explanations in AI," in *Proc. FAT**, 2019, pp. 279–288.
- Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Advances in Neural*

- Information Processing Systems, 2019, pp. 5753–5763
- Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, 2019
- C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, no. 140, pp. 1–67, 2020.
- H. Touvron, T. Lavril, G. Izacard, X. Martinet, M. A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., "LLaMA: Open and efficient foundation language models," arXiv preprint arXiv:2302.13971, 2023.
- H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., "Llama 2: Open foundation and fine-tuned chat models," arXiv preprint arXiv:2307.09288, 2023.
- A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann, et al., "PaLM: Scaling language modeling with pathways," arXiv preprint arXiv:2204.02311, 2022.
- OpenAI, "GPT-4 technical report," arXiv preprint arXiv:2303.08774, 2023.
- Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- A. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proc. ACL*, 2011, pp. 142–150.
- A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *Tech. Rep. CS224N Project Report*, Stanford University, 2009.
- S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 Task 4: Sentiment analysis in Twitter," in *Proc. SemEval*, 2017, pp. 502–518.
- S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "SemEval-2018 Task 1: Affect in tweets," in *Proc. SemEval*, 2018, pp. 1–17.
- J. McAuley and J. Leskovec, "Hidden factors and hidden topics: Understanding rating dimensions with review text," in *Proc. RecSys*, 2013, pp. 165–172.
- D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," in *Proc. ACL*, 2020, pp. 4040–4054.
- K. R. Scherer and H. G. Wallbott, "Evidence for universality and cultural variation of differential emotion response patterning," *Journal of Personality and Social Psychology*, vol. 66, no. 3, p. 310, 1994.
- Nasir, M. S., Khan, H., Qureshi, A., Rafiq, A., & Rasheed, T. (2024). Ethical Aspects In Cyber Security Maintaining Data Integrity and Protection: A Review. *Spectrum of engineering sciences*, 2(3), 420-454.
- Nasir, M. S., Khan, H., Qureshi, A., Rafiq, A., & Rasheed, T. (2024). Ethical Aspects In Cyber Security Maintaining Data Integrity and Protection: A Review. *Spectrum of engineering sciences*, 2(3), 420-454.
- Naveed, A., Khan, H., Imtiaz, Z., Hassan, W., & Fareed, U. (2024). Application and Ethical Aspects of Machine Learning Techniques in Networking: A Review. *Spectrum of engineering sciences*, 2(3), 455-501.
- Nawaz, S., Salman, W., Shahid, U., Khokhar, M. L., Iqbal, M. Z., & Hamid, A. (2024). A Survey on Latest Trends and Technologies of Computer Systems Network. *Spectrum of Engineering Sciences*, 2(4), 85-114.
- Naz, H. Khan, I. Ud Din, A. Ali, and M. Husain, "An Efficient Optimization System for Early Breast

- Cancer Diagnosis based on Internet of Medical Things and Deep Learning", Eng. Technol. Appl. Sci. Res., vol. 14, no. 4, pp. 15957–15962, Aug. 2024
- Niaz, H. U., Qadeer, Q. B. Q., Niaz, H., Mansib, H., Awais, M., & Khan, H. (2025). Artificial Intelligence Assisted Autonomous Unmanned Aerial Vehicles (UAVs) and Aerial drones based on Machine Vision for Enhancing Remote Sensing of Precision crop Health Monitoring. *The Asian Bulletin of Big Data Management*, 5(4), 155-177.
- Niaz, H. U., Qadeer, Q. B. Q., Niaz, H., Mansib, H., Awais, M., & Khan, H. (2025). Artificial Intelligence Assisted Autonomous Unmanned Aerial Vehicles (UAVs) and Aerial drones based on Machine Vision for Enhancing Remote Sensing of Precision crop Health Monitoring. *The Asian Bulletin of Big Data Management*, 5(4), 155-177.
- Noor, A. Ilyas, Z. Javaid, H. Khan, "Framing a Knowledge Domain Visualization on Green Human Resource Management: A Bibliometric Analysis from 2008-2022", *Pakistan Journal of Humanities and Social Sciences.*, vol. 11, no. 4, pp. 4200-4212, Aug. 2023
- Pham, Q.-V., et al. "AI and Big Data for COVID-19 Pandemic." arXiv, 2021.
- Rafay, A., Salman, W., Yahya, G., & Malik, U. (2024). SD Network based on Machine Learning: An Overview of Applications and Solutions. *Spectrum of Engineering Sciences*, 2(4), 150-165.
- Rahman, M. U., Khan, S., Khan, H., Ali, A., & Sarwar, F. (2024). Computational chemistry unveiled: a critical analysis of theoretical coordination chemistry and nanostructured materials. *Chemical Product and Process Modeling*, 19(4), 473-515.
- Raza, A., Khan, H., & Rehman, S. U. (2023). Computational Analysis of Nanomaterials for Energy Storage. *International Journal of Advanced Sciences and Computing*, 143-154.
- Shah, S. Ahmed, K. Saeed, M. Junaid, H. Khan, "Penetration testing active reconnaissance phase–optimized port scanning with nmap tool", In 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE., pp. 1-6, Nov. 2019
- Shaikh, M. S. "Blockchain Applications for Secure Healthcare Big Data." *Healthcare*, MDPI, 2025.
- Sodhro, A. H., et al. "Wearable IoT for Remote Patient Monitoring." *IEEE Sensors Journal*, 2020.
- Sultan, H., Rahman, S. U., Munir, F., Ali, A., Younas, S., & Khan, H. (2025). Institutional dynamics, innovation, and environmental outcomes: a panel NARDL analysis of BRICS nations. *Environment, Development and Sustainability*, 1-43.
- U. Hashmi, S. A. ZeeshanNajam, "Thermal-Aware Real-Time Task Schedulabilty test for Energy and Power System Optimization using Homogeneous Cache Hierarchy of Multi-core Systems", *Journal of Mechanics of Continua and Mathematical Sciences.*, vol. 14, no. 4, pp. 442-452, Mar. 2023
- Ullah, Sana, et al. "Wireless Body Area Networks: Applications and Challenges." *Springer Healthcare IoT*, 2020.
- Farooq, I., Ghafoor, U., Umer, S., Ali, A., Shahid, A. K., & Khan, H. (2025). An Efficient Big Data Security and Privacy in Healthcare for Enhancing Remote Sensing and Monitoring: A Technological Perspective based on ACL for Preserving Big Data Analytics in Cloud. *The Asian Bulletin of Big Data Management*, 5(4), 231-258.
- Waleed, R., Ali, A., Tariq, S., Mustafa, G., Sarwar, H., Saif, S., ... & Uddin, I. (2024). An Efficient Artificial Intelligence (AI) and Internet of Things (IoT's) Based MEAN Stack Technology Applications. *Bulletin of Business and Economics (BBE)*, 13(2), 200-206.
- Xie, J., et al. (2020). 'Blockchain for Electronic Health Records: A Systematic Review.' *IEEE Access*.
- Y. A. Khan, "A GSM based Resource Allocation technique to control Autonomous Robotic Glove for Spinal Cord Implant paralysed Patients using Flex Sensors", *Sukkur IBA Journal of Emerging Technologies.*, vol. 3, no. 2, pp. 13-23, Feb. 2020
- Yousaf, M., Khalid, F., Saleem, M. U., Din, M. U., Shahid, A. K., & Khan, H. (2025). A Deep Learning-Based Enhanced Sentiment Classification and Consistency Analysis of Queries and Results in Search Using Oracle Hybrid Feature Extraction. *Spectrum of Engineering Sciences*, 3(3), 99-121.
- Zaheer, M., Azeem, M. H., Afzal, Z., & Karim, H. (2024). Critical Evaluation of Data Privacy and Security Threats in Federated Learning: Issues and Challenges Related to Privacy and

- Security in IoT. *Spectrum of Engineering Sciences*, 2(5), 458-479.
- Zainab, Khan, H., Din, I. U., Tariq, M. I., Khalid, A., & Naz, A. (2023, May). An Efficient Implementation of an IoT-Based Smart Home Security System. In *International Conference on Computing & Emerging Technologies* (pp. 249-259). Cham: Springer Nature Switzerland.
- Zhang, R., et al. (2023). 'Federated Learning in Medical Imaging.' *Nature Biomedical Engineering*.
- Zhang, Y., Zhang, L., Oki, E., Chawla, N. V., & Kos, A. (2017). IEEE access special section editorial: Big data analytics for smart and connected health. *IEEE Access*, 4, 9906-9909.
- Zhou, J., et al. (2019). 'Privacy-Preserving Data Collection in Healthcare.' *Health Informatics Journal*.



2025 by the authors; The Asian Academy of Business and social science research Ltd Pakistan. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).