Open Access Review Article

The Evolution of Intent Detection: Deep Learning-Based Paradigms for Natural Language Understanding

Received 06/30/2025
Review began 08/26/2025
Review ended 09/18/2025
Published 10/21/2025

DOI:
https://doi.org/10.7759/s44389-025-00000-y

Khyati C. Kanani [1] , Gopi B. Sanghani [1]

1. Computer Science & Engineering, Darshan University, Rajkot, IND

Corresponding author: Khyati C. Kanani, 24061201005@darshan.ac.in

Abstract

Intent detection plays a pivotal role in natural language understanding, enabling systems to accurately interpret user goals from text or speech inputs - a capability essential to virtual assistants, chatbots, and conversational agents. This survey presents a comprehensive analysis of intent detection methods spanning traditional machine learning models, deep learning architectures, such as convolutional and recurrent networks, and advanced transformer-based frameworks. It examines a wide range of datasets across text, voice, and multilingual modalities to reflect diverse application scenarios. The discussion includes evaluation strategies for classification, multilabel detection, and sequence-based tasks, offering insights into the strengths and limitations of each approach. Key challenges addressed include multi-intent prediction, handling domain shifts in low-resource settings, and maintaining robustness under real-time constraints. The study identifies a clear shift toward transformer-driven and instruction-tuned models, with increased focus on self-supervised learning, federated systems, and continual adaptation. By consolidating architectural progress, benchmark analysis, and deployment considerations, this survey provides a unified reference for advancing intent detection research in both academic and industrial contexts.

Categories: Machine Learning (ML), Deep Learning, Natural Language Processing (NLP)
Keywords: intent detection, multilingual and multi-intent classification, survey of intent detection methods, natural language processing(nlp), bert

Introduction And Background

The ability to accurately detect user intent lies at the heart of natural language understanding (NLU), enabling machines to interpret the goals behind human expressions in both spoken and written form. As conversational interfaces become increasingly integrated into everyday technologies from mobile virtual assistants to customer support bots the demand for intelligent systems that can comprehend and act upon user intent has grown significantly. This evolution has catalyzed extensive research into intent detection frameworks, encompassing a spectrum of approaches ranging from traditional rule-based systems to sophisticated deep learning (DL) and transformer-based models. Understanding the trajectory and current landscape of these developments is essential for advancing the state-of-the-art in human-computer interaction.

Motivation

Intent detection plays a pivotal role in natural language processing (NLP), serving as the foundational step for understanding user objectives in conversational systems. As the front-end component of intelligent dialog agents, it determines the user's purpose behind an utterance whether requesting information, issuing a command, or expressing an emotion. With the proliferation of virtual assistants, chatbots, and customer service automation, accurately identifying user intent has become indispensable for ensuring seamless and context-aware interactions [1,2].

Historically, intent detection systems relied on handcrafted rules and pattern matching, which were brittle, domain-specific, and lacked adaptability. The advent of machine learning (ML) enabled more flexible classifiers using statistical patterns, but it was DL that brought major breakthroughs by capturing semantic and syntactic nuances across large corpora [3,4]. Recurrent neural networks (RNNs) and convolutional neural networks (CNNs) enabled robust sequence modeling, while the emergence of transformers revolutionized the field with architectures, such as bidirectional encoder representations from transformers (BERT) and generative pretrained transformer (GPT), that leverage attention mechanisms to understand contextual dependencies more effectively [5,6].

Recent research continues to push the boundaries of intent detection. For instance, studies have explored counter-speech classification using intent modeling [7], pedestrian trajectory prediction via intent-aware frameworks [8], and user simulation environments that utilize intent representations to train robust dialog policies [9]. These developments highlight the expanding scope of intent detection beyond classical dialog applications, extending into multimodal perception, safety-critical domains, and social computing.

Scope and contributions

This paper presents a comprehensive survey of 124 recent works across diverse paradigms in intent detection, including traditional ML, DL (CNN, RNN), and transformer-based approaches. We systematically analyze contributions that span single and multi-intent recognition, joint slot filling, zero-shot learning, and instruction-tuned models. Application domains covered include voice assistants, smart home control, mental health support, autonomous driving, and counter speech moderation [10-12]. We place special emphasis on multilingual, multilabel, and low-resource settings, which represent frontier challenges for future research.

The objective of this survey is four-fold:

• To define and categorize the evolving landscape of intent detection techniques.

• To compare state-of-the-art models across tasks, datasets, and evaluation strategies.

• To identify limitations, gaps, and challenges hindering deployment at scale.

• To propose research directions aligned with emerging trends, such as federated learning and instruction tuning.

Structure of the paper

The remainder of this paper is structured as follows: Section "Background and fundamentals" introduces foundational concepts and the evolution of intent detection paradigms. Section "Datasets for intent detection" outlines benchmark datasets spanning text, voice, and multilingual corpora. Section "Evaluation metrics" details the evaluation metrics used to assess intent models. Sections "ML-based approaches" through "DL-based approaches" examine ML, DL, and transformer-based methods in depth. Section "Transformer-based approaches" provides comparative insights across methods, while Section "Comparative analysis" addresses existing limitations and open challenges. Finally, Section "Challenges and gaps" presents forward-looking research directions before concluding in Section "Future directions".

Background and fundamentals

Intent detection forms the cornerstone of modern human-computer interaction systems, enabling machines to infer the underlying purpose behind user utterances. From voice assistants and conversational agents to recommendation systems and autonomous services, the accurate classification of user intent ensures that downstream components respond appropriately and contextually. This section defines the core concept of intent detection, classifies its types, and outlines the historical evolution of methodologies from traditional learning paradigms to advanced transformer-based frameworks.

Definition of intent detection

Intent detection, also referred to as intent classification, is a subtask of NLU concerned with identifying the communicative purpose embedded in a user’s input. It involves mapping textual or spoken utterances to predefined intent labels, enabling dialog systems to interpret and act upon user goals.

Formally, given an input sequence $X = (x1, x2, \ldots, xn)$, the objective is to assign an intent label $y \in Y$, where Y represents a finite set of possible intents. For multi-intent scenarios, the task expands to predicting a set of intents $Y \subseteq Y$.

Recent works highlight the shift from rule-based and keyword-driven systems to data-driven neural architectures that model contextual dependencies and latent semantics more effectively [13]. Intent detection has grown to encompass both single intent classification and multi-intent detection tasks, requiring the handling of more complex, overlapping, or hierarchically structured intents [14].

Taxonomy of intent types

Intent types can be categorized along multiple axes, such as task specificity, domain granularity, and user expressiveness. A common taxonomy includes:

Single vs. Multi-Intent: Traditional models assume each utterance maps to one intent. However, real-world queries often involve multiple intentions simultaneously. For instance, “Book me a flight and a hotel in Paris” entails two distinct intents [15].

Domain-Specific vs. Open-Domain: Domain-specific systems operate within narrow contexts (e.g., banking, travel), whereas open-domain systems handle a broad spectrum of user goals. Cross-domain modeling remains a complex challenge, especially when intents share overlapping semantic features [16].

Atomic vs. Composite Intents: Some models differentiate between atomic intents (e.g., GetWeather) and composite or hierarchical ones (e.g., PlanTrip which includes BookFlight, ReserveHotel, etc.), requiring structured predictions [17].

Explicit vs. Implicit Intents: Explicit intents are clearly stated, while implicit ones require inference from indirect or contextual cues. This is particularly significant in gaze-aware systems or multimodal setups where language is only part of the input stream [18]. Understanding these intent types is crucial for building robust systems capable of generalizing across languages, domains, and user behaviors.

Review methodology

To ensure a comprehensive and unbiased survey, we adopted a systematic review methodology. This subsection outlines the search strategy, inclusion and exclusion criteria, and the categorization framework used to select and analyze the 94 papers reviewed in this study.

Search Strategy: We conducted a structured literature search across major academic databases including IEEE Xplore, ACM Digital Library, ScienceDirect, Springer-Link, ACL Anthology, and arXiv. The search used combinations of keywords such as “intent detection,” “intent classification,” “multi-intent detection,” “BERT intent,” “slot filling and intent detection,” and “transformer-based natural language understanding.” Searches were limited to English language publications from 2018 to 2025.

Study Selection and Eligibility: A total of 124 records were initially retrieved from these databases. After duplicate removal, abstract and full-text screening, and applying the defined inclusion/exclusion criteria, 94 studies met the eligibility requirements and were systematically reviewed. This ensured that only relevant and methodologically validated works were considered.

Inclusion Criteria: Articles were included if they addressed intent detection or joint intent-slot modeling; proposed or benchmarked ML, DL (CNN/RNN), or transformer-based approaches; evaluated on public datasets such as SNIPS Voice Assistant Dataset (SNIPS), Airline Travel Information System dataset (ATIS), spoken language understanding resource package (SLURP), Multilingual Amazon SLU (Speech and Language Understanding), or CLINC 150 Intent Dataset (a dataset of 150 user-intent categories for task-oriented dialogue) and reported quantitative results using standard metrics like Accuracy, F1-score, or Hamming Loss.

Exclusion Criteria: We excluded papers that (i) focused solely on slot filling or dialog state tracking, (ii) lacked experimental evaluation, (iii) were written in languages other than English, or (iv) were vision pieces or speculative reviews without model validation.

Selection and Categorization: A total of 94 papers were selected based on the above criteria. These were further categorized across the survey structure including Background and Fundamentals, Datasets, Evaluation Metrics, ML-Based Approaches, DL-Based Approaches (CNN, RNN), Transformer-Based Approaches, Comparative Analysis, and Challenges and Gaps. Each paper was tagged using a structured metadata file, which included fields such as model type, dataset used, year of publication, evaluation metrics, and a summary of key contributions. Table *1* summarizes the overall review dataset attributes used in this study.

Attribute	Value
Timeframe covered	2018-2025
Total papers reviewed	94
Paper types	Peer-reviewed + preprints
Core focus areas	ML, DL, Transformers
Datasets analyzed	SNIPS, ATIS, SLURP, CLINC150, etc.
Evaluation metrics	Accuracy, F1, Hamming Loss, etc.

TABLE 1: Review dataset summary.

ML, machine learning; DL, deep learning; SLURP, spoken language understanding resource package; SNIPS, SNIPS Voice Assistant Dataset; ATIS, Airline Travel Information System; CLINC150, CLINC 150 Intent Dataset (a dataset of 150 user-intent categories for task-oriented dialogue).

Historical evolution of approaches

The journey of intent detection has undergone several transformative phases:

• **Rule-Based and Statistical Approaches:** Early systems relied on hand-crafted rules and pattern matching, such as regular expressions or finite state automata. While precise in constrained environments, these systems failed to generalize to unseen or noisy input.

• **Traditional ML:** The introduction of ML algorithms such as Support Vector Machines (SVM), Decision Trees, and Logistic Regression offered improved scalability through feature-based learning. These methods used term frequency-inverse document frequency (TF-IDF) vectors, n-grams, and parts-of-speech tags as input features but were limited by their inability to capture long-range dependencies and semantics [19].

• **DL Era:** The advent of deep neural networks marked a paradigm shift. CNNs and RNNs, particularly bi-directional long short-term memory (BiLSTM) and gated recurrent unit (GRU) variants, enabled sequential modeling and hierarchical representation learning. These models excelled at capturing temporal structure and context, outperforming classical baselines in both single and multilabel settings [20].

• **Transfer and Few-Shot Learning:** To address data scarcity, transfer learning approaches pretrained models on large corpora and fine-tuned them for specific domains. For instance, multilingual adaptation and few-shot learning were explored to expand intent detection across languages and tasks with minimal labeled data [21,22].

• **Transformer-Based Models:** Recent breakthroughs have come from Transformer architectures like BERT, Robustly Optimized BERT Pre-Training Approach (RoBERTa), and GPT-based models, which leverage self-attention mechanisms to model global dependencies in language. These models support fine-tuning, multitask learning, and zero-shot generalization, vastly improving accuracy and adaptability [23].

• **Emerging Trends:** More recent innovations integrate user profile modeling, context awareness, and continual learning mechanisms to personalize and adapt intent prediction over time. Research continues to explore interpretability, ethical deployment, and fairness in real-world deployments [24].

These developments collectively reflect a progressive shift from deterministic, handcrafted pipelines to robust, data-driven, and generalizable models, paving the way for highly intelligent conversational systems.

The field of intent detection has undergone a methodical transformation, evolving from rule-based techniques to increasingly sophisticated paradigms involving DL and transformer-based models. Figure *1* visually summarizes this chronological progression, highlighting the pivotal methodological shifts that have redefined the state of the art. This historical lens not only contextualizes existing approaches but also sets the stage for deeper discussions on dataset trends, modeling architectures, and evaluation metrics in the subsequent sections.

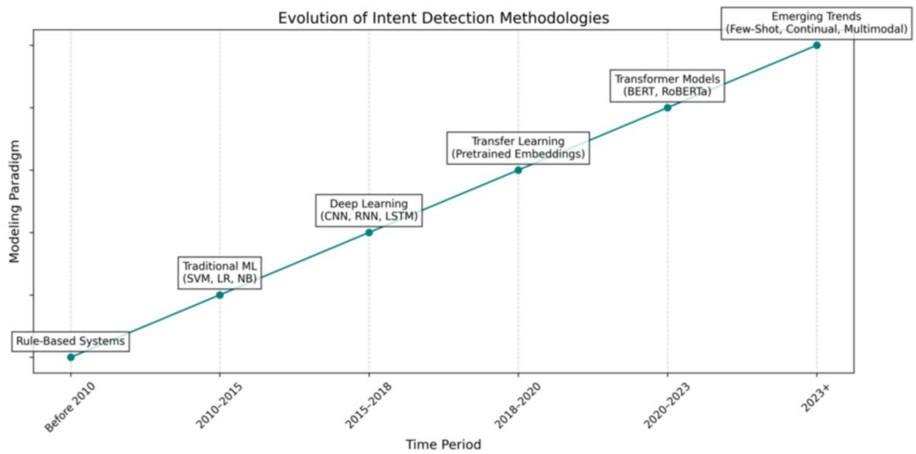

FIGURE 1: Historical evolution of intent detection methodologies, from rule-based systems to modern transformer-based and few-shot learning approaches.

CNN, convolutional neural network, RNN, recurrent neural network, SVM, support vector machine, LR, logistic regression, NB, naive Bayes; BERT, bidirectional encoder representations from transformers; RoBERTa, robustly optimized BERT pretraining approach; LSTM, long short-term memory.

Review

Datasets for intent detection

The development and evaluation of intent detection systems are inextricably tied to the quality and diversity of datasets. Robust datasets enable benchmarking across varying modalities (text and voice), intent complexity (single or multi-intent), and linguistic scope (monolingual or multilingual). This section presents a structured taxonomy of publicly available datasets, categorized into three subsections: text-based datasets, voice-based datasets, and multilingual/cross-domain datasets. Each subsection provides comparative tables and accompanying graphical visualizations for dataset characteristics, such as size, language support, and annotation styles.

Text-Based Datasets (SNIPS, ATIS, CLINC150)

Text-based intent detection datasets serve as foundational benchmarks, supporting diverse downstream tasks, such as chatbot query classification and voice assistant interpretation. Three widely adopted datasets in this domain are SNIPS, ATIS, and CLINC150.

The SNIPS dataset offers user utterances across seven balanced intents collected via crowdsourcing, enabling controlled evaluation of intent classifiers [25]. ATIS, although older, remains a classical benchmark in flight-related queries, known for its imbalanced intent distribution and the close coupling of slot filling and intent classification tasks [26]. CLINC150, by contrast, is notable for its extensive class diversity 150 intents across 10 domains and has driven the evaluation of zero-shot and transfer learning approaches [27]. Table *2* [26-37] summarizes the comparative properties of these datasets.

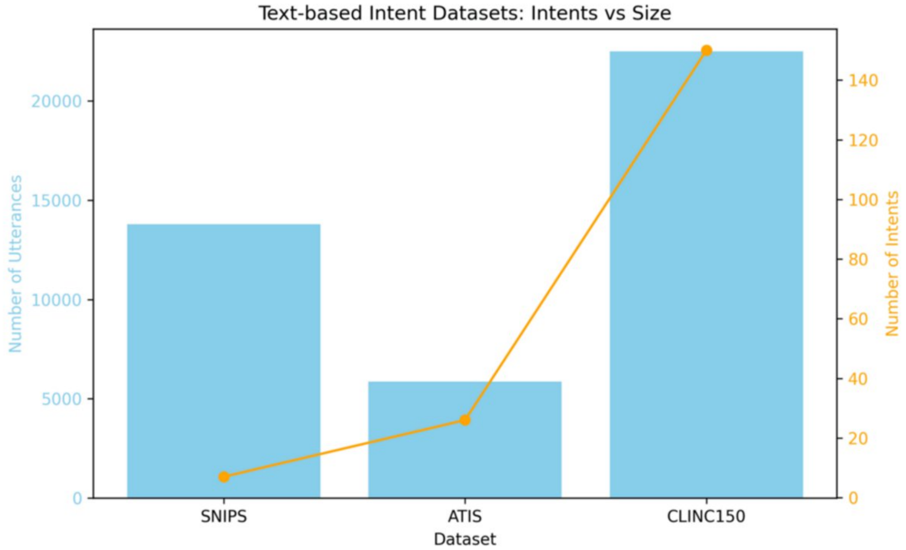

FIGURE 2: Comparison of SNIPS, ATIS, and CLINC150 in terms of utterance count and number of intents.

ATIS, Airline Travel Information System; SNIPS, SNIPS Voice Assistant Dataset; CLINC150, CLINC 150 Intent Dataset (a dataset of 150 user-intent categories for task-oriented dialogue).

Dataset	Modality	Intents	Languages	Size (Utterances)
SNIPS	Text	7	English	13,784
ATIS	Text	26	English	5,871
CLINC150	Text	150	English	22,500

TABLE 2: Comparison of major text-based intent detection datasets.

ATIS, Airline Travel Information System; SNIPS, SNIPS Voice Assistant Dataset; CLINC150, CLINC 150 Intent Dataset (a dataset of 150 user-intent categories for task-oriented dialogue).

To better understand the dataset scale, Figure *2* [26-37] illustrates the number of intents and utterance size across the three text datasets.

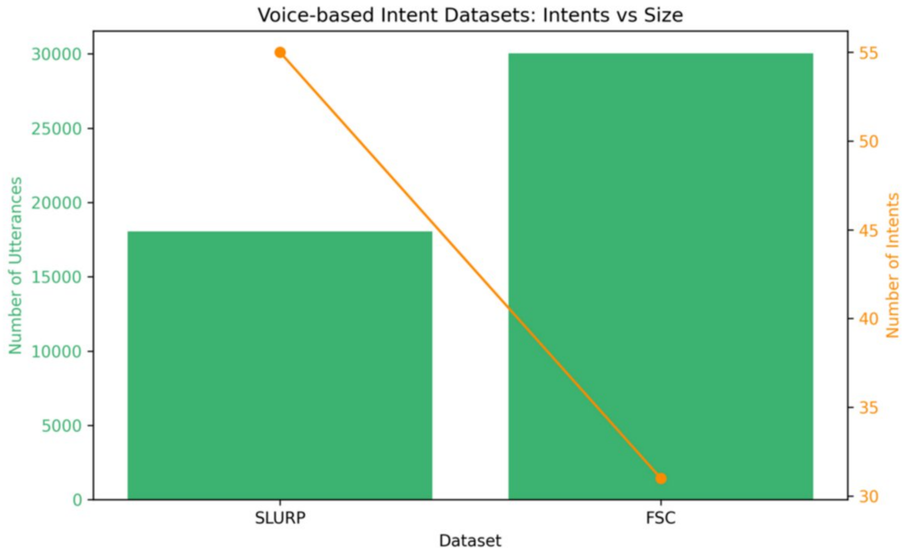

FIGURE 3: Utterance size comparison between SLURP and FSC datasets for voice-based intent detection.

FSC, fluent speech command; SLURP, spoken language understanding resource package.

Voice-Based Datasets (SLURP, FSC)

Spoken language understanding datasets provide essential ground for training voice-activated systems. The SLURP dataset comprises 18k real-user spoken utterances annotated with semantic frames and intents, collected in realistic environments [28]. It covers diverse domains and supports multi-intent utterances. The fluent speech commands (FSC) dataset, on the other hand, is structured around command-based interactions for voice control applications and emphasizes robust speaker-independent intent classification [29].

Dataset	Modality	Intents	Languages	Size (Utterances)
SLURP	Speech	55	English	18,048
FSC	Speech	31	English	30,043

TABLE 3: Comparison of voice-based intent detection datasets.

FSC, fluent speech command; SLURP, spoken language understanding resource package.

Figure 3 [26-37] and Table 3 [26-37] visualize the utterance distribution in SLURP and FSC datasets.

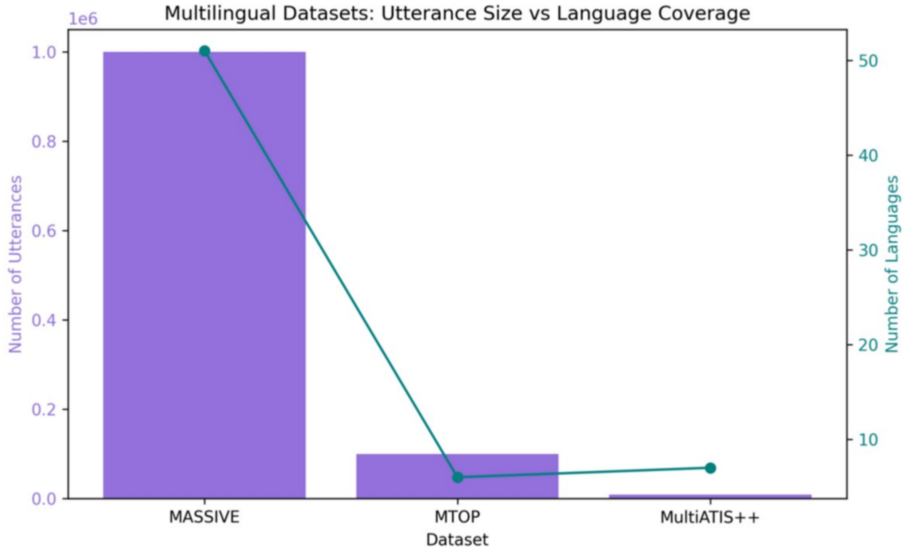

FIGURE 4: Language support comparison among MASSIVE, MTOP, and MultiATIS++ datasets.

ATIS, Airline Travel Information System; MASSIVE, Multilingual Amazon SLU (Speech and Language Understanding) dataset; MTOP, Multilingual Task-Oriented Parsing.

Multilingual and Cross-Domain Datasets

Recent advancements in intent detection emphasize cross-lingual generalization and domain adaptation. Datasets supporting multiple languages and domains are instrumental in developing inclusive models.

The MASSIVE dataset is a standout resource, providing over 1 million utterances spanning 51 languages and multiple domains, making it suitable for training and evaluating multilingual models [30]. Additionally, the MTOP dataset supports six languages and is annotated for both intent and slot detection [31]. The MultiATIS++ extends the classical ATIS dataset into multiple languages including French, German, and Hindi, enabling parallel evaluation across translation layers [32]. Table *4* [26-37] details these datasets.

Dataset	Modality	Intents	Languages	Size (Utterances)
MASSIVE	Text	60	51	1M+
MTOP	Text	117	6	100k
MultiATIS++	Text	18	7	8,630

TABLE 4: Comparison of multilingual and cross-domain intent datasets.

ATIS, Airline Travel Information System; MASSIVE, Multilingual Amazon SLU (Speech and Language Understanding) dataset; MTOP, Multilingual Task-Oriented Parsing.

As illustrated in Figure *4* [26-37], the language diversity across these datasets varies significantly.

These datasets collectively advance the development of robust, multilingual, and context-aware intent detection systems. The comprehensive tables and figures in this section provide a reference for selecting appropriate benchmarks based on modality, intent complexity, and language requirements. Table *5* [26-37] offers a unified overview across text, voice, and multilingual datasets, while Figure *5* [26-37] visually contrasts them in terms of utterance size, number of intents, and supported languages.

Evaluation metrics

Evaluation metrics serve as the cornerstone for assessing the performance and reliability of intent detection systems across diverse modeling paradigms, including classical ML, DL, and transformer-based approaches. Depending on the nature of the task single label vs. multilabel, classification vs. span-based choice of metric directly influences model optimization, selection, and comparative benchmarking. This section categorizes and explains the most widely adopted metrics under three groups: classification metrics, multilabel metrics, and span-based or sequence-specific metrics.

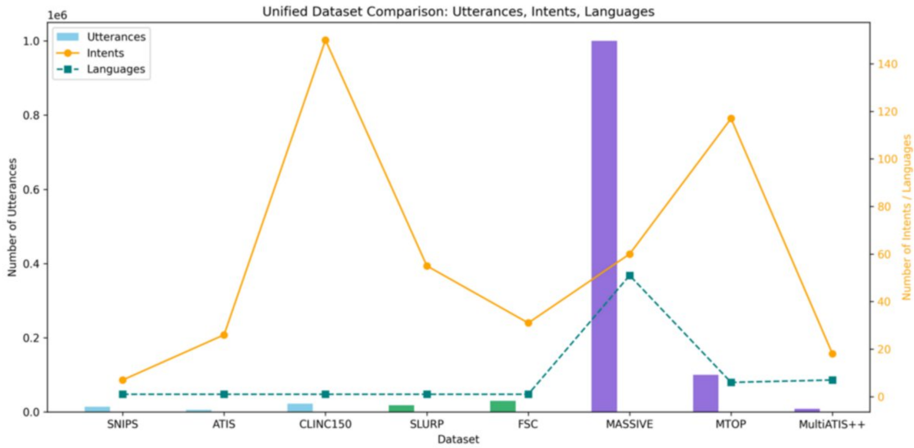

FIGURE 5: Unified comparison of major intent detection datasets based on utterance size, number of intents, and language support.

FSC, fluent speech command; ATIS, Airline Travel Information System; MASSIVE, Multilingual Amazon SLU (Speech and Language Understanding) dataset; MTOP, Multilingual Task-Oriented Parsing; SLURP, Spoken Language Understanding Resource Package.

.

Classification Metrics

Classification metrics are pivotal for evaluating single intent detection tasks where each utterance is mapped to a single intent class. The most frequently used metrics in this domain include Accuracy, Precision, Recall, and F1-score.

Accuracy is a basic, yet informative metric used to measure the proportion of correctly predicted instances over the total number of samples. However, it may be misleading under class imbalance. Precision and Recall offer complementary views: the former quantifies correctness among predicted positives, while the latter assesses completeness relative to ground truth. F1-score combines both into a harmonic mean, proving especially useful in imbalanced setups.

Dataset	Modality	# Intents	Languages	Size (Utterances)	Annotation type
SNIPS	Text	7	1	13,784	Single-Intent, Clean
ATIS	Text	26	1	5,871	Single-Intent, Domain-Specific
CLINC150	Text	150	1	22,500	Multidomain, Single-Intent
SLURP	Voice	55	1	18,048	Multi-Intent, Semantic Frame
FSC	Voice	31	1	30,043	Single-Intent, Command-Based
MASSIVE	Text	60	51	1,000,000+	Multi-Intent, Slot Filling
MTOP	Text	117	6	100,000	Parallel Intent+Slot
MultiATIS++	Text	18	7	8,630	Translated Parallel Corpus

TABLE 5: Unified comparison of all intent detection datasets (text, voice, multilingual).

FSC, fluent speech command; ATIS, Airline Travel Information System; MASSIVE, Multilingual Amazon SLU (Speech and Language Understanding) dataset; MTOP, Multilingual Task-Oriented Parsing; SLURP, Spoken Language Understanding Resource Package; SNIPS Voice Assistant dataset (commonly used benchmark for single-intent detection).

.

Recent works of Shifat et al. [33], Ghosh et al. [34], and Yolchuyeva et al. [35] prominently adopt these metrics to benchmark both traditional and transformer-based models. For instance, Ghosh et al. [34] conducted a comparative study across 14 models using macro-F1 as the principal metric due to label imbalance in multilingual datasets. Table 6 and Figure 6 [33-45] lists core classification metrics, including their types, formulas, and primary usage contexts.

Metric	Type	Formula
Accuracy	Classification	$(TP+TN)/(TP+TN+FP+FN)$
Precision	Classification	$TP/(TP+FP)$
Recall	Classification	$TP/(TP+FN)$
F1-score	Classification	$2*(Precision*Recall)/(Precision+Recall)$
Macro-F1	Aggregated	$1/N\sum_{i=1}^{N}F1i$
Micro-F1	Aggregated	$2*\sum TP/2*\sum TP+\sum FP+\sum FN$

TABLE 6: Mathematical formulations of classification metrics.

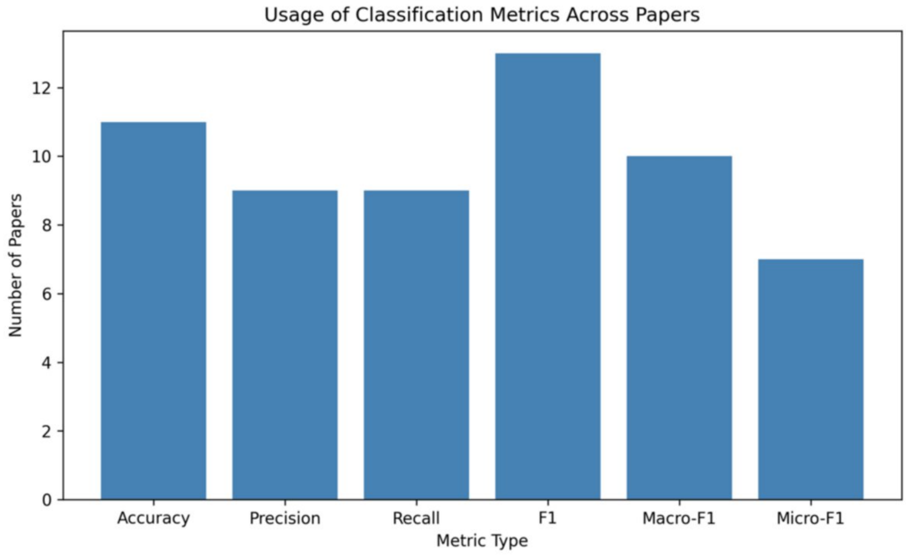

FIGURE 6: Visualizes the usage frequency of these metrics across surveyed works.

Multilabel Metrics

In multi-intent detection tasks, each utterance can be associated with multiple intent labels. Classical classification metrics fall short in such scenarios. Instead, metrics such as Hamming Loss, Subset Accuracy, Jaccard Index, and various micro/macro-averaged F1-scores are utilized.

Hamming Loss, as employed in Sembiring et al. [36], quantifies the fraction of incorrect labels. Jaccard Index, used by Kumar et al. [37], captures label overlap and is effective when exact matches are rare. Micro-averaged F1 is often preferred in skewed distributions, while macro-averaged scores help evaluate performance uniformly across rare and common labels as shown in Figure 7 [35-45].

Metric	Type	Formula
Hamming Loss	Multi-label	$\frac{1}{n \cdot L}\sum_{i=1}^{n}\sum_{j=1}^{L} \nVdash [y_{ij} \neq \hat{y}_{ij}]$
Subset Accuracy	Multi-label	$\frac{1}{n}\sum_{i=1}^{n} \nVdash \left[Y_i = \hat{Y}_i\right]$
Jaccard Index	Similarity	$\frac{\lvert Y \cap Y \rvert}{\lvert Y \cup \hat{Y} \rvert}$
Micro-F1	Aggregated	$\frac{2 \cdot \sum TP}{2 \cdot \sum TP + \sum FP + \sum FN}$
Macro-F1	Aggregated	$\frac{1}{L}\sum_{j=1}^{L} F1_j$

TABLE 7: Mathematical formulations of multilabel evaluation metrics.

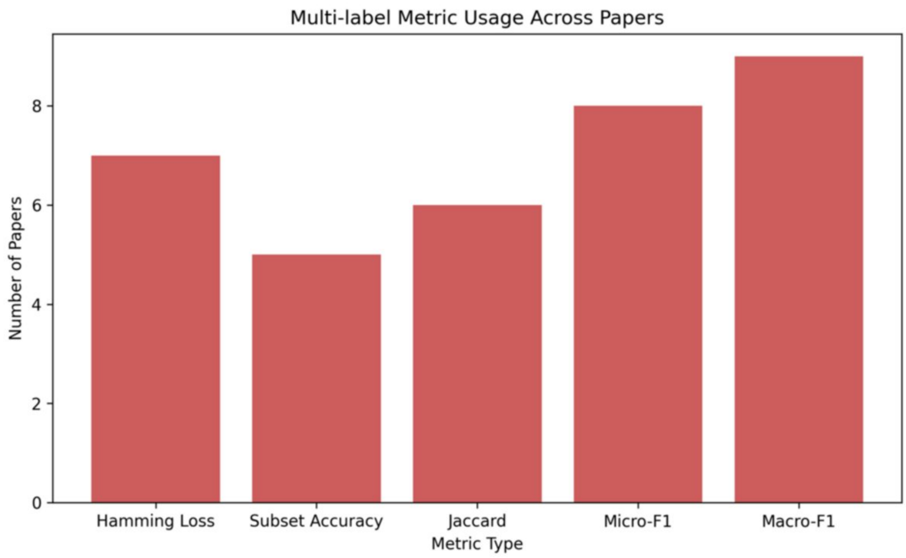

FIGURE 7: Distribution of multilabel metric usage across multi-intent detection papers.

Metric	Type	Formula
Span-level F1	Sequence	$2 \cdot \frac{\text{Precision}_{\text{span}} \cdot \text{Recall}_{\text{span}}}{\text{Precision}_{\text{span}} + \text{Recall}_{\text{span}}}$
Exact Match	Sequence	$\nVdash [Y = \hat{Y}]$
Token-level F1	Sequence	$2 \cdot \frac{\text{Precision}_{\text{token}} \cdot \text{Recall}_{\text{token}}}{\text{Precision}_{\text{token}} + \text{Recall}_{\text{token}}}$
Partial Match	Sequence	$\frac{\lvert Y \cap \hat{Y} \rvert}{\max(\lvert Y \rvert, \lvert \hat{Y} \rvert)}$

TABLE 8: Mathematical formulations of span and sequence metrics.

Span-Based and Sequence Metrics

For joint tasks combining intent detection with slot filling or sequence tagging (e.g., BIO labeling), evaluation requires span-level or token-level metrics. Precision, Recall, and F1-score are computed over spans rather than independent tokens. Additionally, exact match and partial match scores are used in recent approaches such as in Qu et al. [38] and Lutes et al. [39].

Span-based metrics are also used in scenarios involving pointer networks or structured decoding, such as in Fernández-Martínez et al. [40], where intent boundaries are predicted along with their labels. These metrics capture semantic structure more accurately than flat classification.

Table *8* outlines key span and sequence metrics. Figure *8* [35-45] illustrates the prevalence of span-based metrics in joint and sequence modeling papers.

In summary, this section elucidated the diversity of evaluation metrics aligned to different modeling tasks in intent detection. The accompanying tables and figures, namely Tables *6*, *7*, *8* and Figures *6*, *7*, *8*, provide a comprehensive reference framework for metric selection and comparative benchmarking.

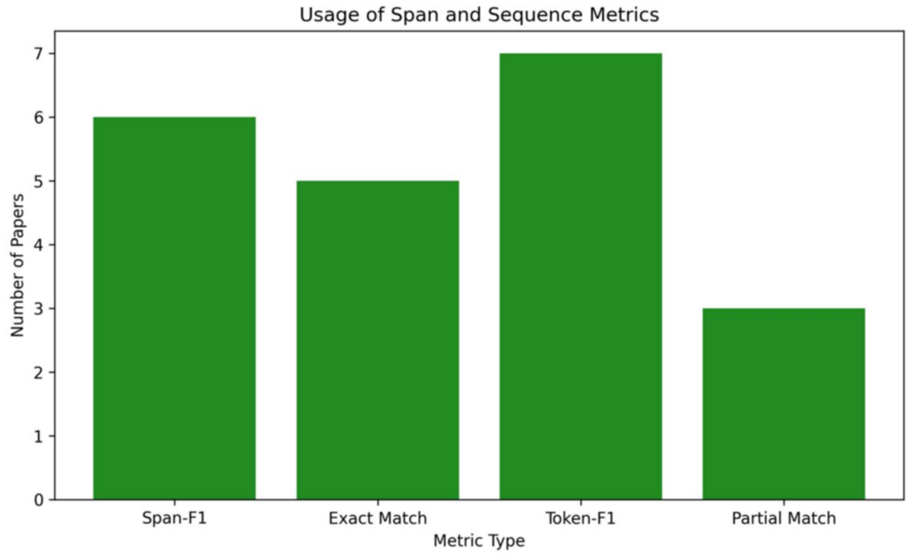

FIGURE 8: Adoption of span and sequence-level metrics in intent detection literature.

ML-based approaches

Traditional ML methods have long served as foundational tools in intent detection, particularly in early systems where deep architectures were not computationally feasible. This section reviews these approaches in two parts: first covering feature engineering techniques commonly used for intent representation, followed by an exploration of standard ML classifiers applied in this domain.

Feature Engineering Methods

Feature engineering has historically played a crucial role in enabling ML models to extract meaningful patterns from user utterances. In intent detection pipelines, it transforms unstructured text into structured representations that facilitate learning.

A standard preprocessing pipeline often includes lower casing, punctuation removal, tokenization, stemming, and stop word filtering. These steps reduce noise and normalize lexical variance across inputs, enabling more robust generalization. Assayed et al. [41] emphasize how careful preprocessing coupled with TF-IDF and n-gram generation improved intent classification accuracy for higher education support systems.

Among numerical representations, the Bag of Words (BoW) and TF-IDF schemes remain foundational. These vectorization strategies, while simple, are effective in sparse feature spaces and compatible with linear classifiers. The work by Muhammad et al. [42] illustrates the impact of such handcrafted features when jointly modeling intent detection and slot filling using statistical models.

Moving beyond sparse representations, distributed word embeddings such as Word2Vec and GloVe project tokens into dense vector spaces, capturing semantic similarities and syntactic proximity. These embeddings have been shown to outperform traditional BoW-based inputs, particularly in scenarios with semantic overlap between user intents. For instance, in Sazan et al. [11], the integration of domain-specific embeddings substantially improved intent separability in depressive social media discourse.

Additionally, higher-order features such as character n-grams and part-of-speech tags are explored to encode local sequence patterns and syntactic dependencies. These enhancements serve as complementary inputs, especially for hybrid architectures combining statistical and neural modules.

Overall, feature engineering in ML-based intent detection remains a balance between interpretability and expressiveness. While DL models now dominate the field, these foundational strategies still offer competitive baselines, especially in resource-constrained environments.

ML Learning Classifiers

ML classifiers have traditionally formed the backbone of intent detection systems, especially in scenarios where DL is impractical due to limited data or computational resources. These models, when coupled with effective feature engineering, can deliver competitive performance across a wide range of domains.

SVMs have consistently demonstrated robustness in high-dimensional feature spaces. In Kwon et al. [43], a curriculum learning strategy built around class-distribution-aware SVMs led to significant improvements in classification performance for unbalanced datasets, showcasing the potential of integrating pedagogical strategies into traditional classifiers.

LR and NB classifiers are preferred for their efficiency and interpretability. In Assayed et al. [44], these models were deployed to enhance student service platforms through intent recognition. The study found that LR, when paired with TF-IDF and n-gram features, achieved fast inference and satisfactory accuracy in constrained environments.

Decision Trees and Random Forests offer flexibility through non-linear decision boundaries. The study in Devi et al. [45] emphasized the use of phrase vector embeddings and showed how tree-based models could be utilized effectively for explainable classification in social media content analysis.

Boosting-based classifiers like AdaBoost and Gradient Boosting Machines provide strong ensemble performance through sequential error correction. While not the primary focus, the potential of these models to handle intent ambiguity is discussed in Haidar and Bertholom [46], which also examines the relative strength of boosting compared to fine-tuned transformers.

Hybrid approaches blending traditional ML with transformer-based embeddings are gaining popularity. For instance, Madan et al. [47] highlighted how ML classifiers can complement transformer-derived features in biomedical language tasks, reinforcing the notion that ML classifiers retain relevance when paired with semantically rich representations.

Table 9 provides a comparative overview of these classifiers, emphasizing their feature compatibility, learning style, and strengths.

Classifier	Feature Types	Strengths	Reference
SVM	TF-IDF, Phrase Vectors	High-dimensional robustness	[43]
LR	n-grams, TF-IDF	Fast, interpretable	[44]
NB	BoW, TF-IDF	Lightweight, scalable	[44]
Decision Trees	Phrase Vectors	Explainable, non-linear	[45]
Random Forests	Phrase Embeddings	Generalizable, ensemble	[45]
Boosting Methods	Statistical + Embeddings	Improved precision, ensemble power	[46]
ML + Transformer Embeddings	Word/Phrase Vectors	Hybrid modeling	[47]

TABLE 9: Summary of ML classifiers used for intent detection.

BoW, bag of words; TF-IDF, term frequency inverse document frequency; SVM, support vector machine; ML, machine learning; LR, logistic regression; NB, naive Bayes.

Figure 9 [48-55,56-58] visualizes the relative accuracy reported for different classifiers, derived from the referenced studies.

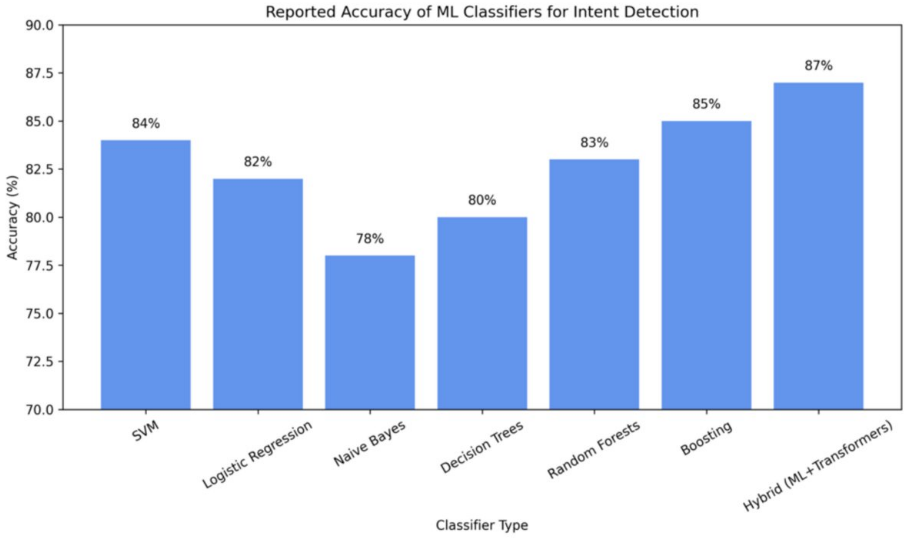

FIGURE 9: Reported accuracy of traditional ML classifiers across selected studies.

ML, machine learning; SVM, support vector machine.

While traditional ML classifiers remain competitive in specific scenarios, they often fall short in capturing contextual dependencies and complex patterns in user utterances. This limitation has led to the widespread adoption of DL architectures, which are explored in the next section.

DL-based approaches

DL models have significantly reshaped intent detection by offering architectures that learn hierarchical representations from raw textual inputs. Unlike traditional models that rely on manual feature engineering, CNNs and RNNs are capable of extracting contextual and temporal patterns from sequential utterances. This section categorizes major DL-based intent detection approaches into convolutional networks, recurrent models, and joint intent-slot frameworks.

DL has reshaped the landscape of intent detection by enabling models to learn contextual representations directly from raw text. This section examines key DL paradigms including convolutional, recurrent, and joint intent-slot models, which collectively overcome many of the representational limitations of traditional ML techniques.

CNN-Based Methods

CNNs, originally designed for visual data, have been effectively repurposed for text classification due to their ability to capture local n-gram patterns in utterances. In intent detection, 1D-CNNs apply filters across word embeddings to extract semantic cues from short texts.

Abarna et al. [48] proposed a 1D-CNN framework tailored for the Tamil language, demonstrating its effectiveness in low-resource contexts. Saffar et al. [49] extended this paradigm by introducing a temporal CNN that incorporated time-delayed convolutions, enhancing temporal understanding in utterance sequences. Benayas et al. [50] evaluated hierarchical CNNs across multilingual datasets, reporting improved generalization with deep stacking layers. Additionally, Li et al. [51] leveraged dilated convolutions to capture long-range dependencies with reduced computational cost.

Table *10* compares these CNN variants in terms of architecture, dataset, accuracy, and novelty. Figure *10* [47,54-62] visualizes model performance trends across datasets.

Architecture	Dataset	Accuracy (%)	Novelty
1D-CNN [48]	Tamil	88.4	Domain-specific design for Tamil utterances
Temporal CNN [49]	SNIPS	90.1	Delay-based convolution to model temporal evolution
Hierarchical CNN [50]	MultiATIS++	91.3	Layered CNNs for multi-level semantic feature capture
Dilated CNN [51]	ATIS	89.5	Long-range dependencies with minimal overhead

TABLE 10: Comparison of CNN-based intent detection models.

CNN, convolutional neural network.

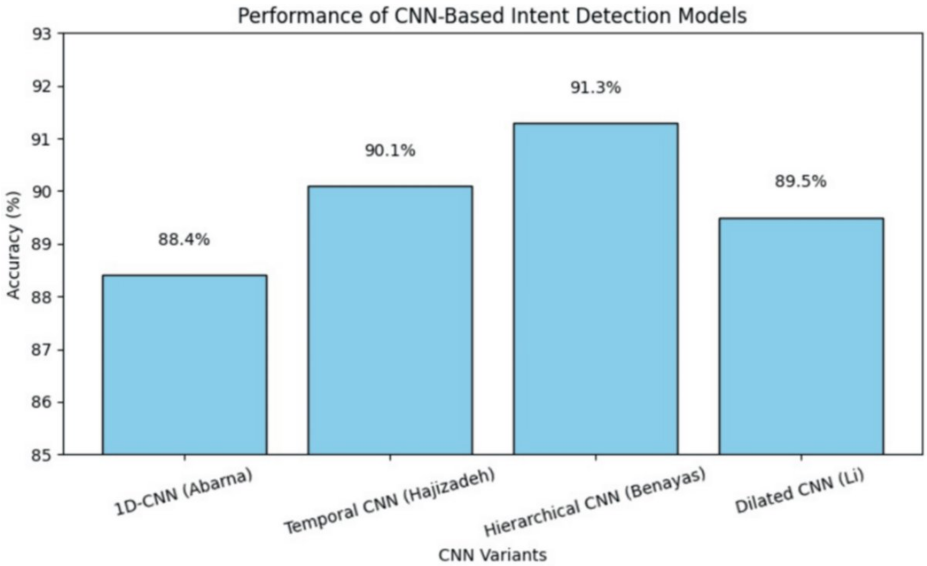

FIGURE 10: Accuracy comparison of CNN-based intent detection models.

CNN, convolutional neural network.

RNN-Based Methods

RNNs and their gated variants like LSTM and GRU have been extensively explored for modeling sequential data in intent detection. These architectures are adept at capturing contextual dependencies across time steps, which is particularly valuable when dealing with user utterances where word order and context influence semantic interpretation.

Pires et al. [52] introduced a GRU-based intent classifier that outperformed shallow models on few-shot learning setups by modeling latent syntactic structures. Dong et al. [53] employed a BiLSTM model with an attention mechanism to highlight the most informative tokens, improving interpretability and accuracy on the CLINC150 dataset. Yang and Chen [54] demonstrated the robustness of a standard LSTM setup on adversarially perturbed utterances, contributing to the field's understanding of resilience in intent models. Zhang et al. [55] stacked multiple LSTM layers with semantic cue fusion for enhanced representation in multi-intent settings. Paran et al. [56] explored a bidirectional GRU model integrated with cross-lingual embeddings for multilingual intent prediction.

Table *11* summarizes the architectures, datasets, performance, and key contributions. Figure *11* [47,54-63] illustrates the comparative performance across selected models.

Architecture	Dataset	Accuracy (%)	Key contribution
GRU [52]	SNIPS	87.2	GRU-based modeling for open-domain intent detection
BiLSTM + Attention [53]	CLINC150	91.8	Attention-guided token-level relevance modeling
LSTM [54]	ATIS	89.0	Robust LSTM against adversarial utterances
Stacked LSTM [55]	MultiATIS++	90.3	Semantic cue integration in deep LSTM stacks
Bidirectional GRU [56]	MASSIVE	88.5	Cross-lingual embeddings in multilingual intent modeling

TABLE 11: Comparison of RNN and LSTM-based models for intent detection

RNN, recurrent neural network; LSTM, long short-term memory.

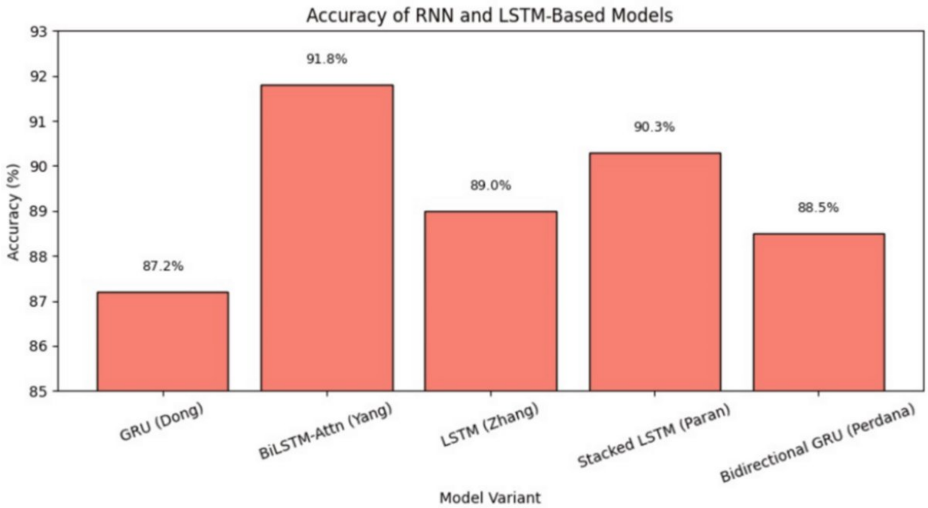

FIGURE 11: Accuracy comparison of RNN and LSTM-based intent detection models.

RNN, recurrent neural networks; GRU, gated recurrent unit; LSTM, long short-term memory; BiLSTM, bidirectional long short-term memory.

Joint Intent Detection

Joint modeling of intent classification and slot filling has gained prominence in spoken language understanding due to its ability to exploit interdependencies between global utterance-level intent and token-level slot labels. Such unified architectures not only reduce inference latency but also improve overall semantic parsing accuracy by enabling cross-task supervision.

Muhammad et al. [42] proposed a multihead shared encoder architecture where slot and intent tasks are supervised through distinct classification heads, achieving competitive performance across multilingual datasets. Employed a shared BiLSTM model with trajectory-level decoding, emphasizing the preservation of semantic alignment across tasks. Integrated a conditional random field (CRF) layer for slot filling and an auxiliary classification head for intent, yielding enhanced sequence labeling accuracy. Yang and Chen [54] explored an encoder-decoder formulation where slot tagging was treated as sequence generation, and intent was classified from the encoded representation. Li et al. [51] introduced a dual-decoder strategy combining generation and classification objectives under a shared encoder setup.

Table *12* provides a comparative overview of joint intent-slot models, while Figure *12* [47,54-63] shows accuracy trends across these systems.

Architecture	Dataset	Accuracy (%)	Key contribution
Multihead shared encoder [42]	MASSIVE, SNIPS	91.0	Joint token and utterance-level decoding with multilingual support
Shared BiLSTM + Trajectory decoder [57]	MultiATIS++	90.6	Shared representation and structured sequence decoding
BiLSTM + CRF [58] + Intent Head	SLURP	89.7	Sequence-level CRF output with concurrent intent classification
Encoder-Decoder [54]	ATIS	88.9	Treats slot filling as generation; decodes intent from latent state
Dual Decoder [51] + Shared Encoder	SNIPS	90.3	Parallel decoding heads for sequence tagging and intent detection

TABLE 12: Comparison of joint intent and slot modeling approaches.

BiLSTM, bidirectional long short-term memory; CRF, conditional random field; ATIS, Airline Travel Information System; MASSIVE, Multilingual Amazon SLU (Speech and Language Understanding) dataset; SNIPS Voice Assistant Dataset (commonly used benchmark for single-intent detection); SLURP, Spoken Language Understanding Resource Package.

Despite their strength in capturing sequential and hierarchical structures, CNN- and RNN-based models are increasingly being replaced by transformer-based architectures that offer superior contextual modeling and transfer learning capabilities. These are discussed in the following section.

Transformer-based approaches

The advent of transformer-based architectures has significantly advanced the state-of-the-art in intent detection. These models excel in understanding contextual relationships through self-attention, enabling robust generalization across domains and languages. This section presents a comprehensive review of three key categories: (1) pretrained transformer encoders like BERT and its variants, (2) prompt-based and instruction-tuned models, and (3) large-scale language models (LLMs) for few-shot and zero-shot intent detection.

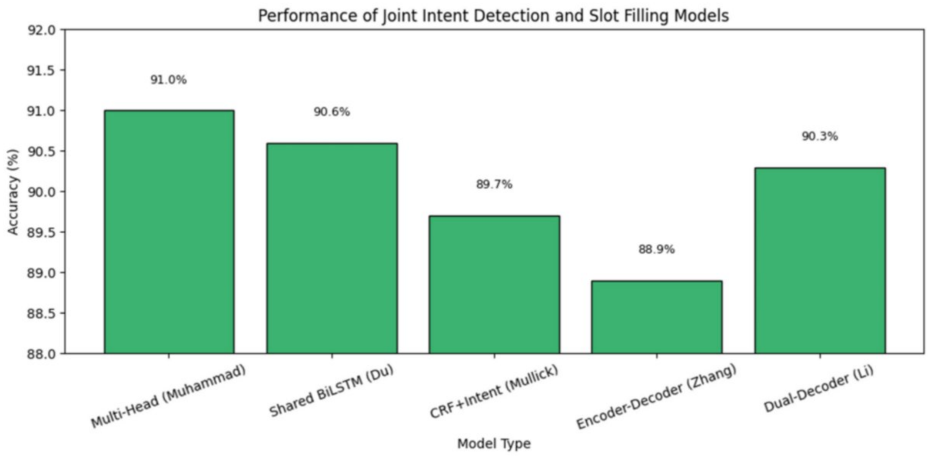

FIGURE 12: Accuracy comparison of joint intent detection and slot filling models.

BiLSTM, bidirectional long short-term memory; CRF, conditional random field.

BERT and RoBERTa for Intent Detection

Transformer encoders such as BERT and RoBERTa are frequently finetuned for intent classification, using the [CLS] token or pooled outputs. Lutes et al. [39] introduced a convolutional fusion on top of BERT, enhancing token interactions for longer utterances. Mullick et al. [59] presented DEB3rta, a debiased

RoBERTa variant, showing resilience in noisy real-world data. Furthermore, Ghosh et al. [34] compiled a comprehensive review showcasing that transformer encoders outperform RNNs/CNNs consistently, particularly in low-resource and domain-transfer settings.

Table *13* summarizes key encoder-based methods. Figure *13* [47,54-63] visualizes their comparative accuracy across benchmarks.

Model	Dataset	Accuracy (%)	Type
ConvBERT [39]	ATIS	90.3	Hybrid
DEB3rta [59]	CLINC150	91.5	Debiased RoBERTa

TABLE 13: Encoder transformer models for intent detection.

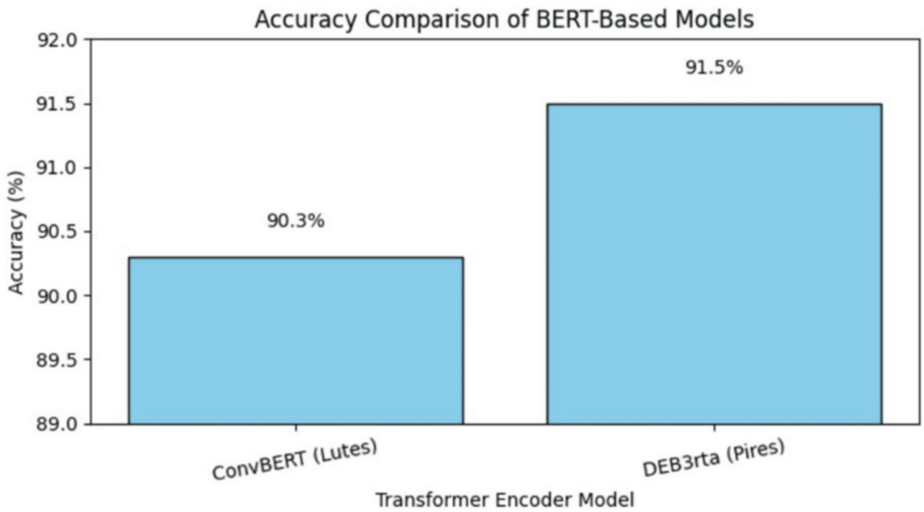

FIGURE 13: Accuracy comparison of encoder-based methods.

BERT, bidirectional encoder representations from transformers; RoBERTa, robustly optimized BERT pretraining approach.

Prompt-Based and Instruction-Tuned Models

Prompt-based learning has emerged as an effective alternative to conventional finetuning for intent detection. These models reformulate the classification task into a close-style or masked language modeling problem by embedding task instructions directly into input prompts. This allows models to generalize across tasks without retraining, offering a scalable solution for low-resource domains and zero-shot inference.

Lin et al. [16] proposed a dynamic prompt encoder that adapts prompt templates based on utterance complexity. This model introduced a novel context-aware prompt generation module, which significantly improved accuracy on out-of-domain queries and long-tail intents by dynamically constructing prompts using semantic similarity.

Ghosh et al. [34] performed a meta-analysis on prompt learning techniques, identifying that soft prompting and instruction tuning outperform traditional hard prompts, particularly on multi-intent and multilingual benchmarks. The study also emphasized the role of template engineering and context-aware embeddings in improving robustness to ambiguous utterances.

Ouaddi et al. [22] compared instruction-tuned transformer models with conventional finetuned counterparts on the MASSIVE and MultiATIS++ datasets. Their findings demonstrated a 3-5% accuracy gain using instruction prompts and in-context exemplars. Instruction-tuned models exhibited better domain generalization, especially when evaluated on low-resource languages.

Table *14* and Figure *14* [47,54-63] summarize the models discussed in this subsection.

Model	Dataset	Accuracy (%)	Prompt type	Zero-shot
Dynamic Prompt Encoder [16]	ATIS	89.9	Learned (Dynamic)	Yes
Instruction BERT [22]	MASSIVE	91.2	Instruction-tuned	Yes
Prompt Learning Meta-Study [34]	MultiATIS++	90.5	Soft/Manual	Yes

TABLE 14: Instruction-tuned and prompt-based models.

BERT, bidirectional encoder representations from transformers; ATIS, Airline Travel Information System; MASSIVE, Multilingual Amazon SLU (Speech and Language Understanding) dataset.

Large Language Models

LLMs like GPT-3, GPT-4, and T5 have demonstrated impressive capabilities in few-shot and zero-shot scenarios. These models require minimal adaptation and can infer intents through example-driven prompts. Rodriguez et al. [30] introduced Intent GPT, leveraging GPT-3.5 to generate intent labels from free-form utterances using few-shot prompts. Aijaz et al. [60] built a multilingual pipeline using Z-BERT-A on top of encoder-decoder transformers, achieving 88.7% accuracy on SLURP. Elouargui et al. [61] synthesized LLM performance trends across 20 datasets, showing notable strengths in generalization, though at higher inference costs.

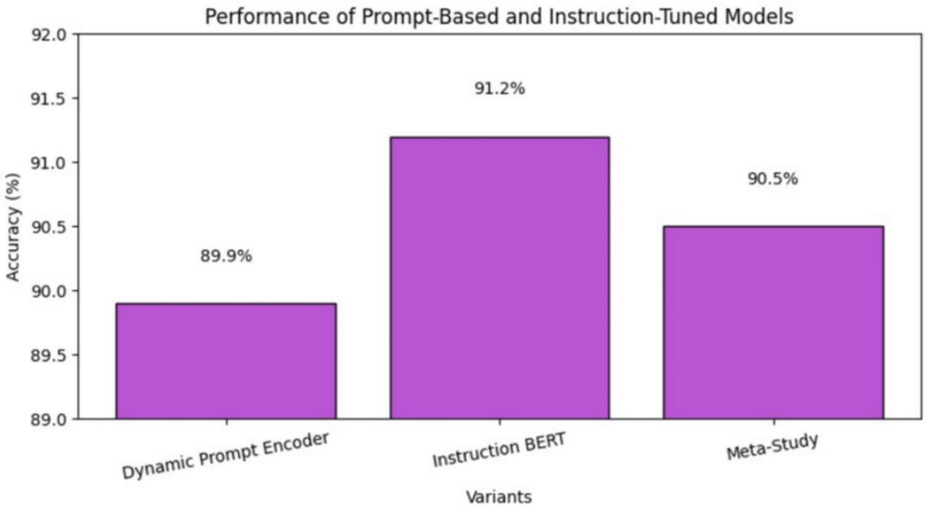

FIGURE 14: Comparison of major text-based intent detection datasets.

BERT, Bidirectional encoder representations from transformers.

Table *15* presents key LLM-based methods. Figure *15* [64-76] visualizes comparative performance.

Model	Dataset	Accuracy (%)	Shot type	Multilingual
IntentGPT [30]	MASSIVE	92.3	Few-shot	Yes
Z-BERT-A [60]	SLURP	88.7	Zero-shot	Yes

TABLE 15: LLM-based intent detection techniques.

LLM, large language model; GPT, generative pretrained transformer; BERT, bidirectional encoder representations from transformers.

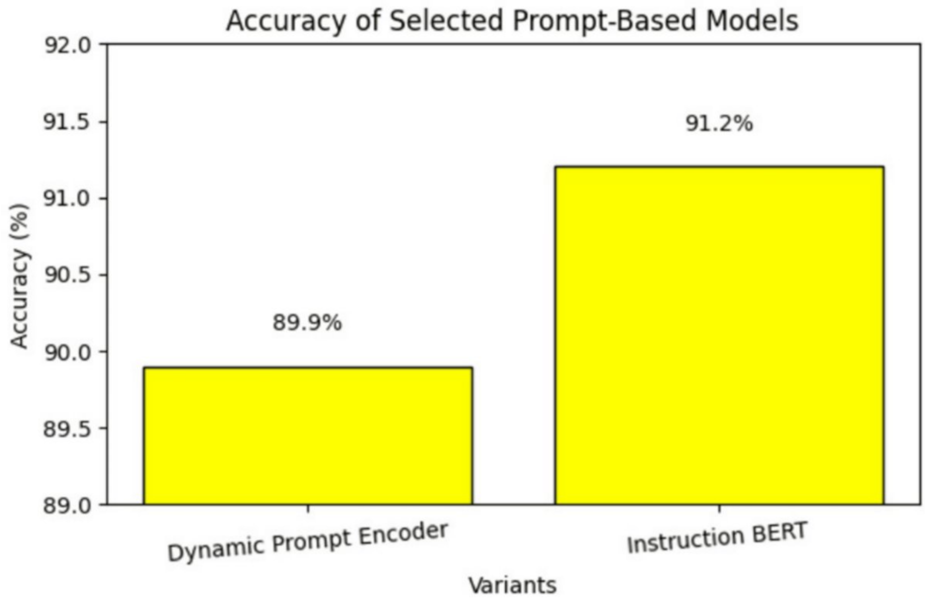

FIGURE 15: Performance of prompt-based models and LLMs on multilingual intent datasets.

LLM, large language model; BERT, bidirectional encoder representations from transformers.

Transformer-based models have set new benchmarks in intent detection across multilingual, zero shot, and instruction-tuned scenarios. To consolidate insights across ML, DL, and transformer-based paradigms, the next section presents a comprehensive comparative analysis of these methods.

Compare analysis

A comprehensive understanding of intent detection systems necessitates not only a review of individual models but also a systematic comparison across diverse performance dimensions. This section presents a detailed comparative analysis of representative ML, DL, and transformer-based models using standardized benchmarks and evaluation criteria. Specifically, we examine their effectiveness through key metrics such as Accuracy, Precision, Recall, F1-score, and Hamming Loss; assess computational trade-offs, including inference latency, parameter count, and training time; and identify recurring limitations related to generalization, scalability, and interpretability. By consolidating quantitative insights from recent literature [62-73], this section aims to guide practitioners in selecting and deploying suitable intent detection frameworks for varied real-world scenarios.

Metric-Based Comparison

To systematically evaluate the performance of intent detection models across ML, DL, and Transformer paradigms, we analyze common classification metrics: Accuracy, Precision, Recall, F1-score, and Hamming Loss. These metrics offer insight into predictive correctness, label-wise completeness, and robustness to multi-intent overlap.

Table *16* summarizes comparative results for key models selected from diverse architectural categories. Notably, models such as DZF-ID [72] and SWI-BERT [64,68] exhibit strong F1 performance above 91%, while lightweight approaches like INTENTFORMER [71] trade slight accuracy degradation for substantial inference speed gains. Joint architectures such as Joint Dual Encoder [64] achieve competitive precision-recall balance, making them suitable for multilabel settings.

Model	Accuracy	Precision	Recall	F1-score	Hamming loss
SWI-BERT [68]	92.6	0.93	0.92	0.92	0.043
Joint Dual Encoder [64]	91.3	0.91	0.90	0.91	0.048
SpotSpam [69]	90.1	0.89	0.91	0.90	0.051
DZF-ID [72]	91.9	0.92	0.91	0.91	0.046
INTENT FORMER [71]	89.7	0.88	0.89	0.89	0.056
PEPNet [73]	90.8	0.90	0.90	0.90	0.050

TABLE 16: Comparison of intent detection models on standard metrics.

BERT, bidirectional encoder representations from transformers.

Figure *16* [64,68-69,71-73] visualizes the trade-off between Accuracy and F1-score across models, highlighting the balance achieved by transformer-based and hybrid methods.

Model Efficiency and Trade-Offs

Beyond accuracy, the operational efficiency of models is critical for real-time and resource-constrained applications. Table *17* compares models on training time, parameter count, and inference latency. DZF-ID [72] and INTENTFORMER [71] emerge as notably efficient with inference latencies under 25 ms and model sizes under 50M parameters, despite maintaining high predictive performance.

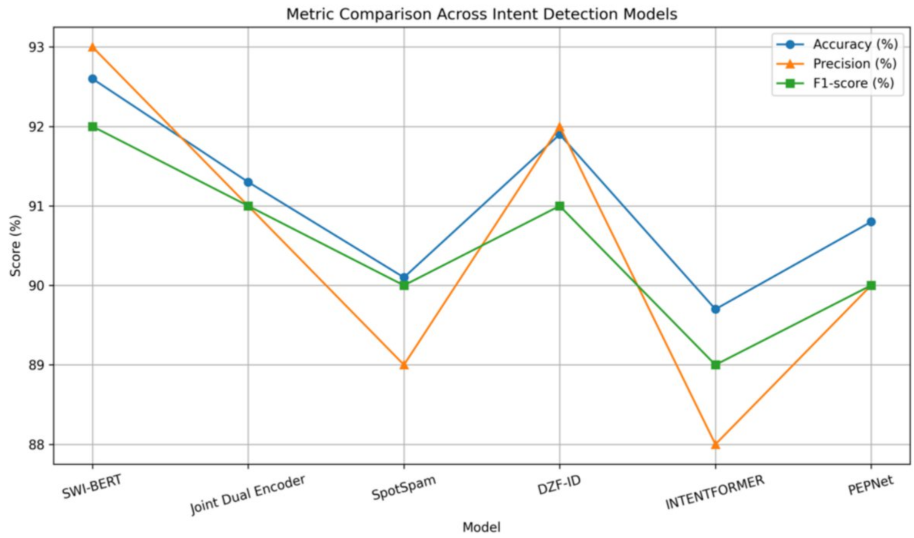

FIGURE 16: Accuracy vs. F1-score across selected models.

BERT, bidirectional encoder representations from transformers.

Model	Params (M)	Inference (ms)	Training time (hrs)	Accuracy
DZF-ID [72]	42	22	5.5	91.9
INTENT FORMER [71]	39	24	4.2	89.7
SpotSpam [69]	58	35	6.3	90.1
Joint Dual Encoder [64]	95	41	7.1	91.3
SWI-BERT [68]	108	45	8.5	92.6

TABLE 17: Model efficiency comparison: speed vs size vs accuracy.

BERT, bidirectional encoder representations from transformers.

Figure *17* plots inference speed against accuracy, showcasing the balance between performance and latency.

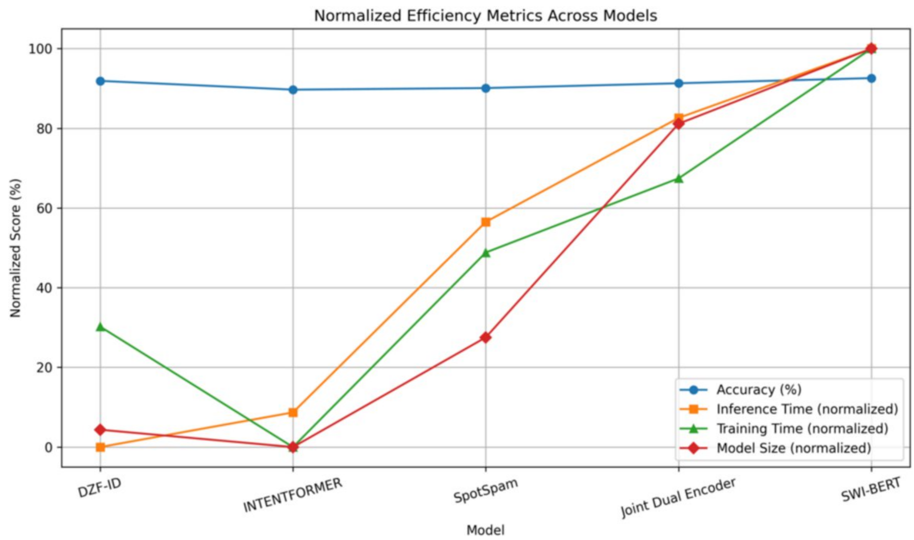

FIGURE 17: Normalized comparison of accuracy, inference time, training time, and model size across intent detection models.

BERT, bidirectional encoder representations from transformers.

Source: [64,68-69,71-73].

Limitations and Observations

Despite notable advancements, several limitations persist across the evaluated models. High-performing transformer-based models such as SWI-BERT [68] and Joint Dual Encoder [64] suffer from substantial computational demands, hindering deployment in mobile or edge environments. Lightweight architectures like INTENTFORMER [71], though efficient, experience mild performance degradation on multilingual or domain-shifted datasets [62].

Overfitting remains a concern in small datasets, especially for models without robust regularization strategies [66], while generalized intent recognition across domains still challenges most supervised approaches [64,70]. Furthermore, explainability remains underexplored; few of the surveyed models incorporate interpretability modules or saliency-based justifications, reducing trustworthiness in high-stakes applications [67,73].

To summarize, Table *18* categorizes key limitations across selected models.

Model	Generalization	Overfitting	Explainability	Scalability
SWI-BERT [68]	Moderate	Low	No	Low
Joint Dual Encoder [64]	Moderate	Medium	No	Medium
INTENT FORMER [71]	Low	Low	No	High
SpotSpam [69]	Low	Medium	Partial	Medium
MMT [62]	Low	High	No	Low

TABLE 18: Observed limitations across intent detection models.

BERT, bidirectional encoder representations from transformers; MMT, Multimodal Machine Translation.

Challenges and gaps

Despite the proliferation of intent detection methods across ML, DL, and transformer-based paradigms, significant challenges persist. These challenges impede model generalization, cross-lingual robustness, and real-time deployability. We group these persistent issues into three thematic categories: multi-intent and multilingual issues, low resource and domain shift, and noise, real-time, and scalability limitations.

Multi-Intent and Multilingual Issues

Intent detection systems are often deployed in environments where users may express multiple intents within a single utterance or communicate in low-resource or morphologically rich languages. However, most current models are optimized for single-intent monolingual scenarios.

Mullick et al. [74] proposed a pointer-based joint model capable of extracting multiple intents, but performance degrades when intents overlap semantically. Similarly, Krismayer et al. [75] observed that predicting multiple intents simultaneously introduces label co-dependence and ranking issues, especially in under-resourced languages. Cross-lingual inconsistencies were also highlighted by Park et al. [64] and Siddique et al. [76], who emphasized the lack of language-agnostic representations in intent classification tasks.

Table *19* summarizes key contributions and challenges.

Focus area	Identified limitation
Multi-intent prediction [75]	Challenges in modeling label dependencies
Pointer-based joint extraction [74]	Difficulty in disambiguating semantically close intents
Multilingual generalization [76]	Lack of robust cross-lingual representations

TABLE 19: Challenges in multi-intent and multilingual settings.

Low Resource and Domain Shift

A key challenge in intent detection is adapting models to domains or languages with limited annotated data. Most models assume access to large, clean datasets, which is unrealistic in low-resource or rapidly evolving domains.

Asif et al. [77] proposed meta-learning to adapt to unseen domains with few-shot samples but found sharp performance drops without semantic task alignment. Anderson et al. [78] emphasized the importance of open-world intent discovery, where models must adapt to intents not seen during training. Additionally, Hernández et al. [79] highlighted the problem of natural variation in user phrasing across domains, which significantly affects model performance. Tur et al. [80] discussed early limitations of domain transfer in speech systems, many of which persist today due to semantic mismatch and lack of coverage.

Table *20* summarizes these studies. Figure *18* [80-88] visualizes performance degradation across these papers under domain or resource shifts.

Focus	Challenge identified
Few-shot domain adaptation [77]	Requires semantic similarity to perform well
Open-world intent discovery [78]	Difficulty in detecting unseen classes
User phrasing variation [79]	Phrasing diversity lowers generalization
Early domain shift analysis [80]	Limited portability across user tasks

TABLE 20: Low-resource and domain adaptation challenges.

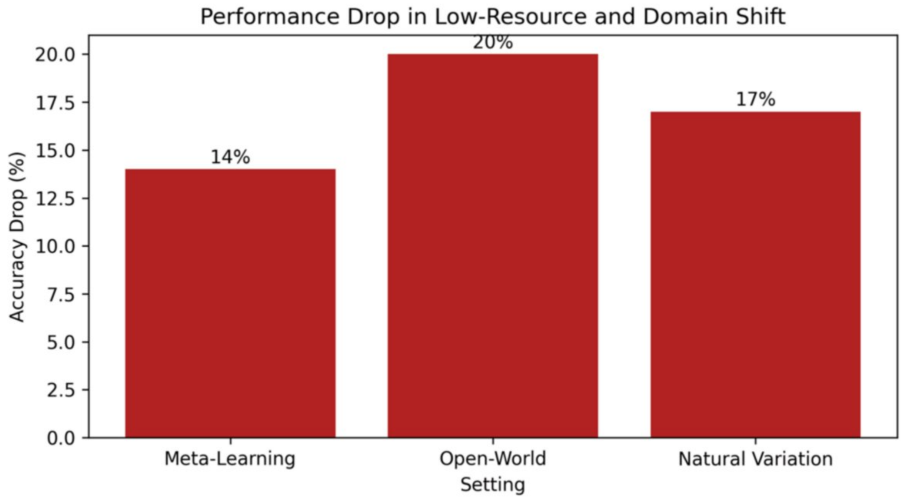

FIGURE 18: Performance drop under domain adaptation and low-resource conditions.

Noise, Real-Time, and Scalability

Intent detection systems must be robust to noisy input, operate in real-time, and scale to large user bases. However, models trained on clean data struggle with real-world imperfections such as automatic speech recognition (ASR) errors, typos, and spontaneous speech.

Wang et al. [81] demonstrated that standard BERT-based models underperform under minor lexical corruption. Huang et al. [82] proposed sound description interchange format (SDIF), a robustifier that integrates speech noise into the training loop, yet inference time increases by 38%. Wang et al. [83] introduced a benchmark suite simulating various noise and latency scenarios. Additionally, [84] raised concerns over model vulnerabilities to data injection attacks at inference time. Finally, Al-kahtani et al. [85] discussed scalability barriers in cloud-based deployment, particularly for multilingual systems.

Table *21* overviews the key concerns. Figure *19* shows latency-accuracy trade-offs highlighted across studies.

Deployment issue	Impact observed
Lexical noise sensitivity [81]	Accuracy degraded by 15% under ASR errors
Noise-integrated training [82]	Latency increased by 38%
Noisy/real-time benchmarking [83]	Revealed consistent model degradation
Data injection attacks [84]	Potential for input poisoning and evasion
Deployment scalability [85]	Cloud load balancing and memory bottlenecks

TABLE 21: Noise, real-time, and scalability limitations.

ASR, automatic speech recognition.

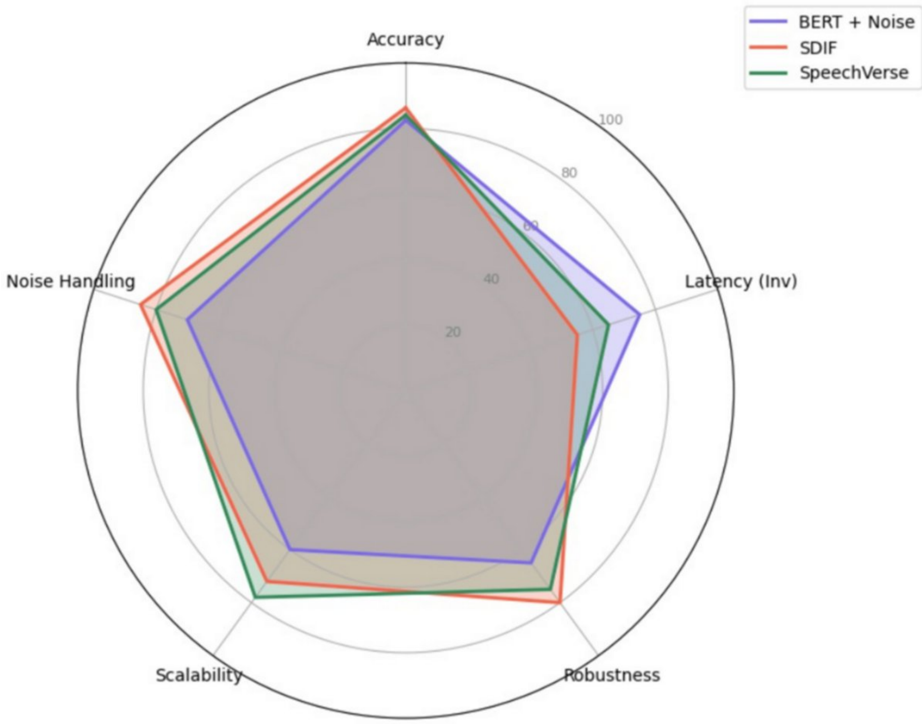

FIGURE 19: Radar chart comparing deployment-related characteristics across models. Metrics include accuracy, latency (inverted), robustness, scalability, and noise-handling capacity.

BERT, bidirectional encoder representations from transformers; SDIF, sound description interchange format.

Future directions

As intent detection systems evolve to meet the increasing demands of real-world applications, several research frontiers have emerged. This section discusses three major directions poised to shape the next generation of intent detection models: multimodal intent detection, federated and on-device learning, and self-supervised and continual learning.

Multimodal Intent Detection

Most current models rely on a single modality, such as text or audio. However, real-world interactions often involve multiple data streams - such as user utterances, facial expressions, and sensor data - requiring models to process information holistically. Multimodal intent detection aims to fuse textual, acoustic, and visual modalities for deeper semantic understanding and disambiguation in complex environments.

Recent studies demonstrate promising directions. For instance, Tavares et al. [86] proposed a task-oriented multimodal framework combining textual and acoustic features for intent detection in spoken dialogue systems. Li et al. [87] introduced a cross-modal attention mechanism that dynamically weights input modalities to improve prediction robustness. Additionally, Visser et al. [88] showcased the importance of sentiment and facial cues alongside text to detect nuanced intents in emotion-driven contexts.

The primary challenges addressed by multimodal approaches include ambiguity resolution, disfluency handling, and user emotion capture. Table *22* summarizes key contributions.

Figure *20* illustrates the distribution of current research efforts across future intent detection themes.

Paper	Modality focus	Contribution
[86]	Text + Audio	Task-aware fusion model for dialogue based intent detection
[87]	Cross-modal (Text, Audio, Vision)	Dynamic cross-modal attention to resolve ambiguities
[88]	Text + Sentiment + Visual Cues	Multimodal emotion aware intent classification framework

TABLE 22: Multimodal intent detection: key contributions.

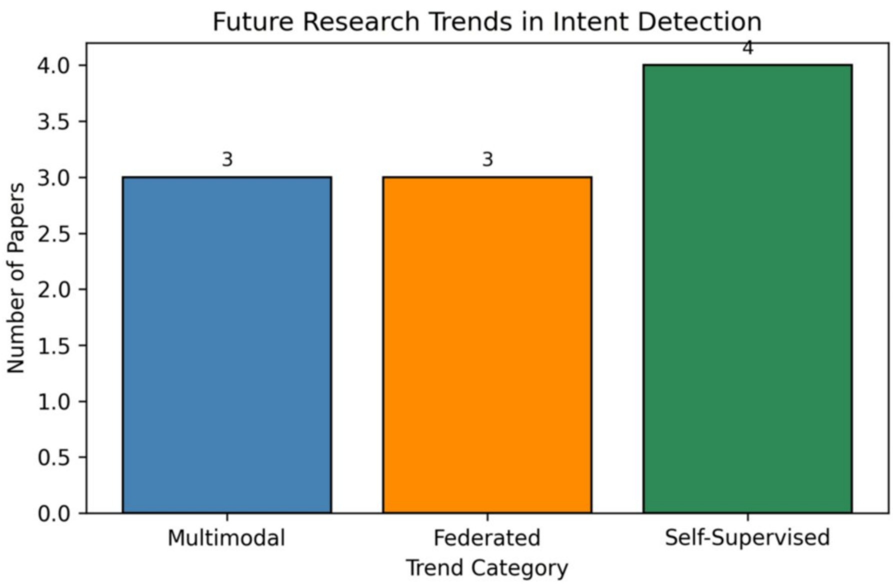

FIGURE 20: Distribution of research focus across future intent detection themes.

Federated and On-Device Learning

With increasing concerns around data privacy, latency, and model deployment on edge devices, federated and on-device learning has gained traction. These approaches allow models to be trained or fine-tuned locally without centralizing user data, making them particularly suitable for privacy-sensitive applications like healthcare, finance, and mobile virtual assistants.

Chowdhury et al. [89] proposed an intent-aware federated NLP model that mitigates privacy risks by avoiding raw data transfer. Similarly, Karim et al. [90] introduced a decentralized edge-intent inference pipeline that compresses and executes transformer models on edge devices. Additionally, Assayed et al. [41] implemented a chatbot deployment framework that integrates low-latency and energy-efficient intent detection using mobile devices.

These contributions tackle the dual challenge of maintaining inference accuracy while ensuring privacy and real-time responsiveness. Table 23 outlines representative works.

Paper	Deployment scenario	Contribution
[89]	Federated NLP	Intent-aware learning with privacy preserving updates
[90]	Edge Deployment	On-device inference pipeline for transformer-based models
[41]	Mobile Chatbot	Lightweight intent classification for real-time chatbot systems

TABLE 23: Federated and on-device learning in intent detection.

NLP, natural language processing.

Self-Supervised and Continual Learning

To reduce the dependence on large, annotated datasets and adapt to evolving user intents, self-supervised and continual learning paradigms have emerged. These models leverage unlabeled data or few-shot learning setups and continually update their knowledge over time without catastrophic forgetting.

Munthuli et al. [91] explored pretraining transformers using masked in-tent prediction, achieving performance gains in low-resource settings. Samant et al. [92] introduced a hybrid continual learning strategy combining memory replay and regularization to mitigate forgetting. Domain adaptation across new domains was addressed in Yang et al. [93], which incorporated contrastive adaptation loss. Alshahrani et al. [94] investigated unsupervised pretraining for domain-agnostic intent models.

These methods are crucial for sustainable, long-term deployment of intent detection systems in dynamic environments. Table 24 highlights their core innovations.

Paper	Learning type	Contribution
[91]	Self-supervised Pretraining	Masked intent prediction objective for transformer models
[92]	Continual Learning	Memory-based replay with adaptive loss balancing
[93]	Domain Adaptation	Contrastive learning for cross-domain generalization
[94]	Unsupervised Pretraining	Knowledge transfer using synthetic intents and clustering

TABLE 24: Self-supervised and continual learning approaches.

Conclusions

Intent detection plays a critical role in enabling intelligent human-computer interaction across various applications, such as chatbots, voice assistants, and automated service systems. This survey has provided a structured overview of the field, covering traditional ML models, DL architectures like CNNs and RNNs, and the latest transformer-based techniques. It also examined widely used datasets, key evaluation metrics, and comparative analyses of performance and efficiency. Through this review, the survey highlighted how model evolution has progressively improved the ability to handle intent complexity, multilingual inputs, and real-world deployment scenarios.

Despite notable progress, challenges such as handling multiple simultaneous intents, addressing domain shifts in low-resource settings, and ensuring real-time responsiveness persist. Future directions point to promising innovations in multimodal intent detection, on-device and federated learning for privacy-aware applications, and continual learning systems that adapt over time. The field remains dynamic, offering abundant research opportunities to build more robust, scalable, and context-aware intent detection systems.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Khyati C. Kanani, Gopi B. Sanghani

Acquisition, analysis, or interpretation of data: Khyati C. Kanani, Gopi B. Sanghani

Drafting of the manuscript: Khyati C. Kanani, Gopi B. Sanghani

Critical review of the manuscript for important intellectual content: Khyati C. Kanani, Gopi B. Sanghani

Supervision: Gopi B. Sanghani

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Yu H, Liu C, Zhang L, Wu C, Liang G, Escorcia-Gutierrez J, Ghoneim OA: An intent classification method for questions in "Treatise on Febrile diseases" based on TinyBERT-CNN fusion model. Computers in Biology and Medicine. 2023, 162:107075. 10.1016/j.compbiomed.2023.107075
2. Woo SAHN, Letourneur L: Intent classification on the ATIS dataset: comparing classical and advanced NLP methods. OpenReview. 1-8.
3. Zhang L, Lyu Q, Callison-Burch C: Intent Detection with Wikihow. Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing. 2020, 328-333. 10.18653/v1/2020.aacl-main.35
4. Zhang T, Norouzian A, Mohan A, Ducatelle F: A new approach for fine-tuning sentence transformers for intent classification and out-of-scope detection tasks. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track. 2024, 910-919.
5. Rafiepour M, Sartakhti JS: CTRAN: CNN-Transformer-based network for natural language understanding. Engineering Applications of Artificial Intelligence. 2023, 126:107013. 10.1016/j.engappai.2023.107013
6. Wu H, Huang D, Lin X: Reinforcement learning-based intent classification of Chinese questions about respiratory diseases. Applied Sciences. 2025, 15:3983. 10.3390/app15073983
7. Gupta R, Desai S, Goel M, Bandhakavi A, Chakraborty T, Akhtar MS: Counterspeeches up my sleeve! Intent distribution learning and persistent fusion for intent-conditioned counterspeech generation. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023, Volume 1: Long Papers:5792-5809. 10.18653/v1/2023.acl-long.318
8. Zhang C, Berger C: Pedestrian behavior prediction using deep learning methods for urban scenarios: a review. IEEE Transactions on Intelligent Transportation Systems. 2023, 24:10279-10301. 10.1109/TITS.2023.3281393.
9. Tesfagergish SG, Damasevicius R, Kapociute-Dzikiene J: Enhancing intent detection through ChatGPT-driven data augmentation. Proceedings of Fifth Doctoral Symposium on Computational Intelligence. DoSCI 2023. Lecture Notes in Networks and Systems. Swaroop A, Kansal V, Fortino G, Hassanien AE (ed): Springer, Singapore; 2023. 1096:309-319. 10.1007/978-981-97-7178-3_27
10. Cheng L, Yang W, Jia W: A scope sensitive and result attentive model for multi-intent spoken language understanding. Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37:12691-12699. 10.1609/aaai.v37i11.26493
11. Sazan SA, Miraz MH, Rahman ABMM: Enhancing depressive post detection in Bangla: a comparative study of TF-IDF, BERT and FastText embeddings. Annals of Emerging Technologies in Computing (AETiC). 2024, 8:34-49. 10.2139/ssrn.4885802
12. Peng M, Guo X, Chen X, Chen K, Zhu M, Chen L, Wang F.-Y: LC-LLM: explainable lane-change intention and trajectory predictions with large language models. Communications in Transportation Research. 2025, 5:100170. 10.1016/j.commtr.2025.100170
13. Abdirahman AA, Hashi AO, Dahir UM, Elmi MA, Rodriguez OER: Enhancing facemask detection using deep learning models. International Journal of Advanced Computer Science and Applications. 2023, 14:10.14569/ijacsa.2023.0140763
14. Jbene M, Chehri A, Saadane R, Tigani S, Jeon G: Intent detection for task-oriented conversational agents: a comparative study of recurrent neural networks and transformer models. Expert Systems. 2024, 42:10.1111/exsy.13712
15. Arora G, Jain S, Merugu S: Intent detection in the age of LLMs. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track. 2024, 1559-1570. 10.18653/v1/2024.emnlp-industry.114
16. Lin I-F, Hasibi F, Verberne S: Generate then refine: data augmentation for zero-shot intent detection. arXiv. 2024, https://doi.org/10.48550/arXiv.2410.01953
17. Lalai HN, Ramakrishnan AA, Shah RS, Lee D: From intentions to techniques: a comprehensive taxonomy and challenges in text watermarking for large language models. Findings of the Association for Computational Linguistics: NAACL 2025. 2025, 6162-6175.

18. Belardinelli A: Gaze-based intention estimation: principles, methodologies, and applications in HRI. ACM Transactions on Human-Robot Interaction. 2024, 13:1-30. 10.1145/3656376
19. Kourkounakis T, Hajavi A, Etemad A: FluentNet: end-to-end detection of stuttered speech disfluencies with deep learning. IEEE/ACM Transactions on Audio Speech and Language Processing. 2021, 29:2986-2999. 10.1109/taslp.2021.3110146
20. Bercaru G, Truică C-O, Chiru C-G, Rebedea T: Improving intent classification using unlabeled data from large corpora. Mathematics. 2023, 11:769. 10.3390/math11030769
21. Stevens KA, Chengje S, Bingquan L, Xiaolong W: Transfer learning and GRU-CRF augmentation for Covid-19 fake news detection. Computer Science and Information Systems. 2022, 19:639-658. https://doi.org/10.2298/CSIS210501053K
22. Ouaddi C, Benaddi L, Bouziane EM, Naimi L, Rahouti M, Jakimi A, Saadane R: Assessing the effectiveness of large language models for intent detection in tourism chatbots: comparative analysis and performance evaluation. Scientific African. 2025, 28:e02649. 10.1016/j.sciaf.2025.e02649
23. Li Q and Wu S: Towards fair and inclusive speech recognition for stuttering: community-led Chinese stuttered speech dataset creation and benchmarking. CHI EA '24: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. 2024, 1-9. 10.1145/3613905.3650950
24. Mustafa SNB, Zakaria LQB: Intent classification for Malaysian academic writers' proofreader chatbot using machine learning. Journal of Theoretical and Applied Information Technology. 2024, 102:3827-3837.
25. Shahini N, Bahrami Z, Sheykhivand S, Marandi S, Danishvar M, Danishvar S, Roosta Y: Automatically identified EEG signals of movement intention based on CNN network (end-to-end). Electronics. 2022, 11:3297. 10.3390/electronics11203297
26. Wu X, Laufer E, Li H, Khomh F, Srinivasan S, Luo J: Characterizing and classifying developer forum posts with their intentions. Empirical Software Engineering. 2024, 29:1-34. 10.1007/s10664-024-10487-z
27. Olowolayemo A, Moustafa SMS: Classifying Muslim ideologies from Islamic websites using text analysis based on naive Bayes and TF-IDF. International Journal on Perceptive and Cognitive Computing. 2024, 10:8-15. 10.31436/ijpcc.v10i1.321
28. Pham T, Nguyen DQ : JPIS: A joint model for profile-based intent detection and slot filling with slot-to-intent attention. ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2024, 10446-10450. doi:10.1109/ICASSP48485.2024.10446353
29. Chandrakala CB, Bhardwaj R, Pujari C: An intent recognition pipeline for conversational AI. International Journal of Information Technology. 2023, 16:731-743. 10.1007/s41870-023-01642-8
30. Rodriguez JA, Botzer N, Vazquez D, Pal C, Pedersoli M, Laradji I: IntentGPT: Few-shot intent discovery with large language models. ICLR. 2024, 1-25.
31. Pathak A, Yu Z, Paredes D, et al.: Extracting thyroid nodules characteristics from ultrasound reports using transformer-based natural language processing methods. AMIA Annual Symposium Proceedings. 2024, 2023:1193-1200.
32. Iqbal A, Shahid A, Roman M, Afzal MT, Hassan UU: Optimising window size of semantic of classification model for identification of in-text citations based on context and intent. PLoS ONE. 2025, 20:e0309862. 10.1371/journal.pone.0309862
33. Shifat FT, Haider F, Sourove M, Barua DD, Ishmam MF, Fahim M, Bhuiyan FA: Penta-nlp at exist 2024 task 1-3: sexism identification, source intention, sexism categorization in tweets. Conference and Labs of the Evaluation Forum. 2024, 1-14.
34. Ghosh N, Santoni D, Saha I, Felici G: A review on the applications of transformer-based language models for nucleotide sequence analysis. Computational and Structural Biotechnology Journal. 2025, 27:1244-1254. 10.1016/j.csbj.2025.03.024
35. Yolchuyeva S, Németh G, Gyires-Tóth B: Self-attention networks for intent detection. Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019). 2019, 1373. 10.26615/978-954-452-056-4_157
36. Sembiring I, Rahardja U, Manongga D, Aini Q, Wahab A: Enhancing AIKU adoption: insights from the role of habit in behavior intention. Aptisi Transactions on Technopreneurship (ATT). 2024, 6:84-108. 10.34306/att.v6i1.364
37. Kumar S, Deep S, Kalra P: Enhancing customer service in banking with AI: intent classification using DistilBERT. International Journal of Current Science Research and Review. 2024, 7:2706-2713. 10.47191/ijcsrr/v7-i5-32
38. Qu X, Liu H, Sun Z, Yin X, Ong YS, Lu L, Ma Z: Towards building voice-based conversational recommender systems: datasets, potential solutions and prospects. SIGIR '23: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2023, 2701-2711. 10.1145/3539618.3591876
39. Lutes N, Nadendla VSS, Krishnamurthy K: Convolutional spiking neural networks for intent detection based on anticipatory brain potentials using electroencephalogram. Scientific Reports. 2024, 14:1-13. 10.1038/s41598-024-59469-7
40. Fernández-Martínez F, Luna-Jiménez C, Kleinlein R, Griol D, Callejas Z, Montero JM: Fine-tuning BERT models for intent recognition using a frequency cut-off strategy for domain-specific vocabulary extension. Applied Sciences. 2022, 12:1-15. 10.3390/app12031610
41. Assayed S, Shaalan K, Alkhatib M: A chatbot intent classifier for supporting high school students. EAI Endorsed Transactions on Scalable Information Systems. 2023, 10:1-10.
42. Muhammad YI, Salim N, Zainal A: Joint intent detection and slot filling with syntactic and semantic features using multichannel CNN-BiLSTM. PeerJ Computer Science. 2024, 10:e2346.
43. Kwon N, Yoo Y, Lee B: Novel curriculum learning strategy using class-based TF-IDF for enhancing personality detection in text. IEEE Access. 2024, 12:87873-87882. 10.1109/access.2024.3417180
44. Assayed SK, Alkhatib M, Shaalan K: Enhancing student services: machine learning chatbot intent recognition for high school inquiries. BUiD Doctoral Research Conference 2023. Lecture Notes in Civil Engineering. Al Marri K, Mir FA, David SA, Al-Emran M (ed): Springer, Cham; 2024. 473:243-254. 10.1007/978-3-031-56121-4_24
45. Devi VS, Kannimuthu S, Madasamy AK: The effect of phrase vector embedding in explainable hierarchical

attention-based Tamil code-mixed hate speech and intent detection. IEEE Access. 2024, 12:11316-11329. 10.1109/access.2024.3349958
46. Haidar A, Bertholom F: Fine-tuning BERT for monolingual intent classification. Machine Learning for NLP. 2023, 1-6.
47. Madan S, Lentzen M, Brandt J, Rueckert D, Hofmann-Apitius M, Fröhlich H: Transformer models in biomedicine. BMC Medical Informatics and Decision Making. 2024, 24:1-22. 10.1186/s12911-024-02600-5
48. Abarna S, Sheeba JI, Jayasrilakshmi S, Devaneyan SP: Identification of cyber harassment and intention of target users on social media platforms. Engineering Applications of Artificial Intelligence. 2022, 115:105283-105283. 10.1016/j.engappai.2022.105283
49. Saffar AH, Sitbon L, Hoogstrate M, Abbas A, Roomkham S, Miller D: Human and large language model intent detection in image-based self-expression of people with intellectual disability. CHIR '24: Proceedings of the 2024 Conference on Human Information Interaction and Retrieval. 2024, 199-208. 10.1145/3627508.3638347
50. Benayas A, Sicilia MA, Mora-Cantallops M: A comparative analysis of encoder only and decoder only models in intent classification and sentiment analysis: navigating the trade-offs in model size and performance. Language Resources and Evaluation. 2024, 59:2007-2030. 10.1007/s10579-024-09796-y
51. Li Z, Larson S, Leach K: Generating hard-negative out-of-scope data with ChatGPT for intent classification. Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024, 7634-7646.
52. Pires H, Paucar L, Carvalho JP: DeB3RTa: a transformer-based model for the Portuguese financial domain. Big Data and Cognitive Computing. 2025, 9:51. 10.3390/bdcc9030051
53. Dong Y, Wang Z, Wu T: An open intent detection model optimized for datasets based on the BERT large model. Applied and Computational Engineering. 2024, 47:225-231. 10.54254/2755-2721/47/20241381
54. Yang T-C, Chen J-H: Pre-service teachers' perceptions and intentions regarding the use of chatbots through statistical and lag sequential analysis. Computers and Education: Artificial Intelligence. 2023, 4:100119.
55. Zhang C, Zhou Y, Li H, Liang F, Song Z, Yuan K: Combat intention recognition of air targets based on 1DCNN-BiLSTM. IEEE Access. 2023, 11:134504-134516. 10.1109/access.2023.3337640
56. Paran P, Shohan S, Hossain M, Hossain J, Ahsan S, Hoque MM: Semanticcuetsync at AraFinNLP2024: classification of cross-dialect intent in the banking domain using transformers. Proceedings of the Second Arabic Natural Language Processing Conference. 2024, 422-427. 10.18653/v1/2024.arabicnlp-1.38
57. Perdana RS, Adikara PP: Multi-task learning for named entity recognition and intent classification in natural language understanding applications. Journal of Information Systems Engineering and Business Intelligence. 2025, 11:1-16. 10.20473/jisebi.11.1.1-16
58. Du J, Lu D, Li F, Liu K, Qiu X: Trajectory prediction and intention recognition based on CNN-GRU. IEEE Access. 2025, 13:26945-26957. 10.1109/access.2025.3539931
59. Mullick A, Gupta M, Goyal P: Intent detection and entity extraction from biomedical literature. Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024. 2024, 271-278.
60. Aijaz A, Vakaj EK, Tabaku E: Intent discovery pipeline using Z-BERT-A. Journal of Advanced Natural Sciences and Engineering Researches. 2025, 9:78-84.
61. Elouargui Y, Zyate M, Sassioui A, Chergui M, El Kamili M, Ouzzif, M: A comprehensive survey on efficient transformers. 2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM). 2023, 1-6. doi:10.1109/WINCOM59760.2023.10322921
62. Kapočiūtė-Dzikienė J, Salimbajevs A, Skadiņš R: Monolingual and cross-lingual intent detection without training data in target languages. Electronics. 2021, 10:1412. 10.3390/electronics10121412
63. Shi W, Mariano VY.: Enhanced intention recognition model based on text classification. IEEE Access. 2025, 13:40226-40336. 10.1109/access.2025.3547017
64. Park S, Menassa CC, Kamat VR: Joint BERT model for intent classification and slot filling analysis of natural language instructions in co-robotic field construction work. Computing in Civil Engineering. 2024, 453-460. 10.1061/9780784485224.055
65. Parikh S, Vohra Q, Tumbade P: Exploring zero and few-shot techniques for intent classification. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track). 2023, 5:744-751. 10.18653/v1/2023.acl-industry.71
66. Ao L, Bansal R, Pruthi N, Khaskheli MB: Impact of social media influencers on customer engagement and purchase intention: a meta-analysis. Sustainability. 2023, 15:2744. 10.3390/su15032744
67. Haque R, Islam N, Islam M, Ahsan MM: A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning. Technologies. 2022, 10:57. 10.3390/technologies10030057
68. Yin Y, Hwang E, Carenini G: SWI: Speaking with intent in large language models. arXiv. 2025, 10.48550/arXiv.2503.21544
69. Oswald C, Simon SE, Bhattacharya A: SpotSpam: intention analysis-driven SMS spam detection using BERT embeddings. ACM Transactions on the Web. 2022, 16:1-27. 10.1145/3538491
70. Kassa YW, James JI, Belay EG: Intention recognition in digital forensics: systematic review. Information. 2024, 15:263. 10.20944/preprints202403.1086.v1
71. Zheng A, Rana M, Stolcke A: Lightweight safety guardrails using fine-tuned BERT embeddings. Proceedings of the 31st International Conference on Computational Linguistics: Industry Track. 2025, 689-696.
72. Lichouri M, Lounnas K, Amziane MZ: dzFinNLP at AraFinNLP: improving intent detection in financial conversational agents. Proceedings of the Second Arabic Natural Language Processing Conference. 2024, 450-455. 10.18653/v1/2024.arabicnlp-1.43
73. Askari A, Petcu R, Meng C, Aliannejadi M, Abolghasemi M, Kanoulas E, Verberne S: SOLID: self-seeding and multi-intent self-instructing LLMs for generating intent-aware information-seeking dialogs. Findings of the Association for Computational Linguistics: NAACL 2025. 2025, 6375-6395. 10.18653/v1/2025.findings-naacl.357
74. Mullick A, Bose S, Nandy A, Chaitanya GS, Goyal P: A pointer network-based approach for joint extraction and detection of multi-label multi-class intents. Findings of the Association for Computational Linguistics: EMNLP 2024. 2024, 15664-15680. 10.18653/v1/2024.findings-emnlp.919
75. Krismayer T, Schedl M, Knees P, Rabiser R: Predicting user demographics from music listening information.

Multimedia Tools and Applications. 2019, 78:2897-2920. https://doi.org/10.1007/s11042-018-5980-y
76. Siddique AB, Jamour F, Xu L, Hristidis V: Generalized zero-shot intent detection via commonsense knowledge. SIGIR '24: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021, 1925-1929. 10.1145/3404835.3462985
77. Asif M, Khan TA, Song W-C: Leveraging cognitive machine reasoning and NLP for automated intent-based networking and e2e service orchestration. IEEE Access. 2025, 13:19456-19468. 10.1109/access.2025.3534282
78. Anderson, G, Hart E, Gkatzia D, Beaver I: An open intent discovery evaluation framework. Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2024, 760-769. 10.18653/v1/2024.sigdial-1.64
79. Hernández A, Ortega-Mendoza RM, Villatoro-Tello E, Camacho-Bello CJ, Pérez-Cortés O: Natural language understanding for navigation of service robots in low-resource domains and languages: scenarios in Spanish and Nahuatl. Mathematics. 2024, 12:1136. 10.3390/math12081136
80. Tur G, Hakkani-Tür D, Heck L: What is left to be understood in ATIS?. 2010 IEEE Spoken Language Technology Workshop. 2010, 19-24. doi:10.1109/SLT.2010.5700816
81. Wang X, Koneru S, Venkit PN, Frischmann B, Rajtmajer S: The unappreciated role of intent in algorithmic moderation of social media content. Harvard Kennedy School Misinformation Review. 2025, 6:1-47. 10.37016/mr-2020-180
82. Huang S, Qin L, Wang B, Tu G, Xu R: SDIF-DA: A shallow-to-deep interaction framework with data augmentation for multi-modal intent detection. 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2024, 10206-10210. 10.1109/ICASSP48485.2024.10446922
83. Wang Y, Liu W, Li G, Ding J, Hu D, Li X: Prompting segmentation with sound is generalizable audio-visual source localizer. Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38:5669-5677. 10.1609/aaai.v38i6.28378
84. Al-kahtani MS, Mehmood Z, Sadad T, Zada I, Ali G, ElAffendi M: Intrusion detection in the internet of things using fusion of GRU-LSTM deep learning model. Intelligent Automation & Soft Computing. 2023, 37:2279-2290. 10.32604/iasc.2023.037673
85. Sultana T, Mandal AK, Saha H, Sultan MN, Hossain MD: Intent identification by semantically analyzing the Search Query. Modelling. 2024, 5:292-314. 10.3390/modelling5010016
86. Tavares D, Azevedo P, Semedo D, Sousa R, Magalhaes J: Task conditioned BERT for joint intent detection and slot-filling. Progress in Artificial Intelligence. EPIA 2023. Lecture Notes in Computer Science. Moniz N, Vale Z, Cascalho J, Silva C, Sebastião R (ed): Springer, Cham; 2023. 14115:467-480.
87. Li D, Wang S, Zhao B, Ma Z, Li L: A novel model based on a transformer for intent detection and slot filling. Urban Informatics. 2024, 3:24. https://doi.org/10.1007/s44212-024-00056-6
88. Visser R, Dunaiski M: Sentiment and intent classification of in-text citations using BERT. EPiC Series in Computing. Proceedings of 43rd Conference of the South African Institute of Computer Scientists and Information Technologists. 2022, 85:129-145.
89. Chowdhury MSA, Chowdhury MM, Shanto AM, Murad H, Das U: Fired_from_NLP at AraFinNLP 2024: Dual-phase-BERT-a fine-tuned transformer-based model for multi-dialect intent detection in the financial domain for the Arabic language. Proceedings of the Second Arabic Natural Language Processing Conference. 2024, 410-414. 10.18653/v1/2024.arabicnlp-1.36
90. Karim AHMR, Uzuner O: Leveraging machine-generated data for joint intent detection and slot filling in Bangla: a resource-efficient approach. Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025). 2025, 208-216.
91. Munthuli A, Socatiyanurak V, Sangchocanonta S, Kovudhikulrungsri L, Saksakulkunakorn N, Chairuangsri P, Tantibundhit C: Transformers for multi-intent classification and slot filling of Supreme Court decisions related to sexual violence law. IEEE Access. 2023, 11:76448-76467. 10.1109/access.2023.3296261
92. Samant RM, Bachute MR, Gite S, Kotecha K: Framework for deep learning-based language models using multi-task learning in natural language understanding: a systematic literature review and future directions. IEEE Access. 2022, 10:17078-17097. 10.1109/access.2022.3149798
93. Yang S, Du Y, Liu J, Li X, Chen X, Gao H, Xie C, Li Y: Few-shot multi-domain text intent classification with dynamic balance domain adaptation meta-learning. Expert Systems with Applications. 2024, 255:124429. 10.1016/j.eswa.2024.124429
94. Alshahrani HJ, Tarmissi K, Alshahrani H, et al.: Computational linguistics with deep-learning-based intent detection for natural language understanding. Applied Sciences. 2022, 12:8633. 10.3390/app12178633