International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Advancing Text Classification: A Systematic Review of Few-Shot Learning Approaches

Amani Aljehani [1], Syed Hamid Hasan [2] and Usman Ali Khan [3]

[1] *Department of Information Systems, College of Computer Science and Information Technology, King Abdulaziz University, Jeddah, KSA.*
[2] *Department of Information Systems, College of Computer Science and Information Technology, King Abdulaziz University, Jeddah, KSA.*
[3] *Department of Information Systems, College of Computer Science and Information Technology, King Abdulaziz University, Jeddah, KSA.*

E-mail address: aaljehani0217@stu.kau.edu.sa, shhasan@kau.edu.sa, ukhan@kau.edu.sa

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Few-shot learning, a specialized branch of machine learning, tackles the challenge of constructing accurate models with minimal labeled data. This is particularly pertinent in text classification, where annotated samples are often scarce, especially in niche domains or certain languages. Our survey offers an updated synthesis of the latest developments in few-shot learning for text classification, delving into core techniques such as metric-based, model-based, and optimization-based approaches, and their suitability for textual data. We pay special attention to transfer learning and pre-trained language models, which have demonstrated exceptional capabilities in comprehending and categorizing text with few examples. Additionally, our review extends to the exploration of few-shot learning in Arabic text classification, including both datasets and existing research efforts. We evaluated 32 studies that met our inclusion criteria, summarizing benchmarks and datasets, discussing few-shot learning's real-world impacts, and suggesting future research avenues. Our survey aims to provide a thorough groundwork for those at the nexus of few-shot learning and text classification, with an added focus on Arabic text, emphasizing the creation of versatile models that can effectively learn from limited data and sustain high performance, while also identifying key challenges in applying Few-Shot Learning (FSL), including data sparsity, domain specificity, and language constraints, necessitating innovative solutions for robust model adaptation and generalization across diverse textual domains.

Keywords: Few-shot learning, Text classification, Transfer learning, Machine Learning, Pre-trained Language Models

1. Introduction

Text classification is a crucial task in natural language processing (NLP), which assigns predefined labels to textual content, facilitating the organization and retrieval of information. The process is integral to a variety of linguistic analyses, including syntactic categorization and semantic analysis, where the methods range from simple Bayesian classifiers to complex deep learning models [1, 2]. The evolution of these classification methodologies, from Bayesian approaches to sophisticated neural networks, mirrors the rapid advancement in the field of NLP, which addresses increasingly complex and nuanced language data [3]. This progress exemplifies how text classification has adapted to the ever-growing demands of processing and understanding large volumes of text in diverse applications. The utility of these techniques extends beyond mere data categorization; they are essential in extracting meaningful patterns from unstructured text, a common form of data in the digital age [4]. Recent advances have seen the integration of context-aware algorithms, which significantly improve the accuracy and granularity of classification tasks [5].

As the volume of digital documents grows across various domains, text classification becomes increasingly crucial, leveraging NLP to efficiently manage, sift, and extract value from an expanding repository of textual content over time. These methodologies underpin a multitude of applications that are central to the current information-driven society. For instance, this technology automates document sorting in digital libraries [6], improves targeted advertising by classifying user interests [7], and improves customer service with automated

E-mail:author's email

response systems [8]. Furthermore, it ensures the delivery of relevant content on social networks and sentiment analyses [9, 10, 11, 12], helps in the processing of clinical documents in healthcare [13, 14, 15], filters spam emails for cybersecurity [16, 17, 18, 19], and analyzes customer feedback in business intelligence [20, 21, 22]. These varied applications highlight the indispensable role of text classification across multiple domains, demonstrating its impact from digital media to healthcare and business.

Traditional text classification systems have been dependent on large volumes of annotated data to achieve high accuracy [23]. Furthermore, the static nature of these systems often means they struggle to adapt to the evolving nature of language, making them less effective as new terms and usage patterns emerge [24].

To mitigate these challenges, few-shot learning presents a promising avenue for training models effectively with a limited amount of labeled data [25, 26]. This is particularly advantageous for applications dealing with rare events or emerging topics where sufficient training data may not be available [27]. Moreover, few-shot learning models have the flexibility to adapt quickly to new tasks, which is essential in the rapidly changing landscape of digital communication [28].

The primary objective of this study is to conduct a systematic and comprehensive literature review of the few-shot learning approach in text classification. This involves rigorously defining the scope and objectives, with a focus on examining the application and effectiveness of few-shot learning techniques specifically in the context of text classification tasks. Moreover, this entails a detailed examination of various methodologies and comparative analyses of few-shot learning against traditional models. Additionally, the review seeks to identify pivotal research questions that explore the efficacy, adaptability, and future prospects of few-shot learning approaches in text classification. The purpose is to provide a comprehensive understanding of the current landscape and potential growth areas in this field, providing a clear and structured narrative for both academic and practical applications.

2. LITERATURE REVIEWS AND BACKGROUND

This section explores various studies on few-shot learning and text classification. It begins with an analysis of related works and surveys. Then, it reviews approaches to few-shot learning, primarily in English, with some attention to multilingual models. The section also covers contributions to text classification, including research on Arabic text, offering a comprehensive research overview.

A. Surveys

In the field of text classification, encompassing both few-shot learning and general methodologies, a wealth of literature reviews have provided in-depth insight. For instance, the systematic review by Ge *et al.*[29] specifically focuses on few-shot learning in the context of medical text, providing an in-depth exploration of this niche area. Similaryl, Le Glaz *et al.*[30] present a systematic review on the use of machine learning and natural language processing in mental health, highlighting the advancements and challenges in this research area. Yang's *et al.* [31] provides a comprehensive overview of the advancements in few-shot learning within the realm of natural language processing. Ignaczak *et al.*'s paper (2021) provides a systematic literature review on text mining in cybersecurity, offering a comprehensive overview of the state of the field. Complementing this, Minaee *et al.* [32] present a broad overview of deep learning-based text classification methods, covering a wide range of deep learning architectures and their applications in text classification. Similarly, Li *et al.*in 2020 [33] and two years later (2022) [34] provide extensive surveys that trace the evolution of text classification methods from shallow learning techniques to the latest deep learning approaches. Additionally, Wu *et al.* [35] review various deep learning methods specifically in the context of text classification, evaluating their effectiveness and comparing different approaches. Brauwers *et al.* [36] as well as Mabrouk *et al.* [37] propose a surveys on aspect-based sentiment classification. These collective works offer a comprehensive view of the current trends, methodologies, and challenges in the field of text classification.

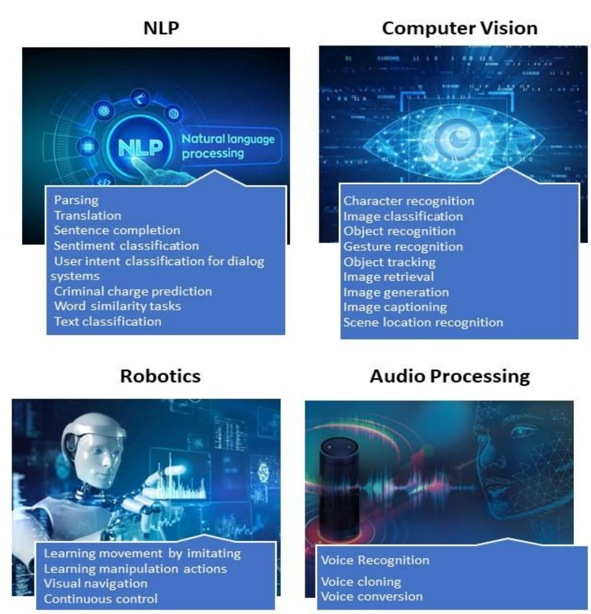

Figure 1. Multiple use cases of Few-Shot learning across diverse domains

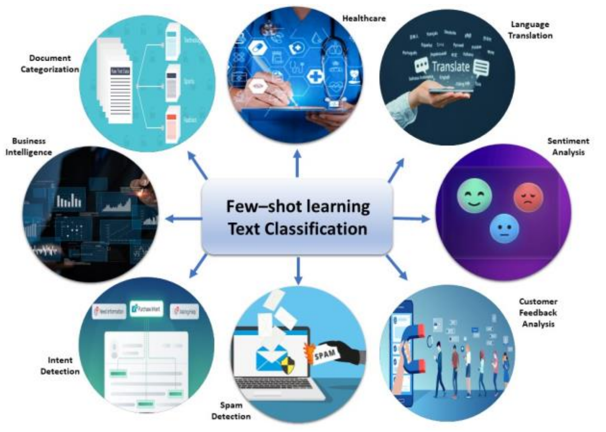

Figure 2. Explored use cases of Few-Shot Learning in the context of text classification domain

B. Few-Shot Learning in NLP 2.3. English Language

Globally, the predominant language, both in speaker numbers and as the main conduit for scholarly inquiry, is English. Such prevalence has cultivated a rich tapestry of literature, aspects of which we will delve into and illuminate in the following sections of our document. Ye *et al.* (2021) [38] introduce a challenge for assessing cross-task generalization capabilities in NLP. Bragg *et al.* (2021) [39] contribute to unifying the evaluation framework for few-shot NLP. Hou *et al.* (2020) [40] present "FewJoint," a benchmark for few-shot learning in joint language understanding. Lastly, Perez *et al.* (2021) [41] delve into the true potential of few-shot learning with language models in their study. Liu *et al.* (2023) [9] and Alqahtani *et al.*(2022) [10] have demonstrated the effectiveness of few-shot learning models in sentiment classification.

C. Multi-Language

Researchers in few-shot learning with multilingual language models focus on creating systems that can process and comprehend several languages simultaneously [42, 43]. In the context of few-shot learning for Arabic language tasks, research is sparse; however, one notable contribution is the introduction of JASMINE [44], a suite of Arabic autoregressive Transformer language models, which demonstrates significant advancements in few-shot learning across a broad spectrum of NLP tasks.

D. Text Classifications

In the expansive landscape of literature, numerous works have delved into text classification. Sentiment analysis is a notable example, with a wealth of studies dedicated to exploring its intricacies [45, 46, 47]. In the realm of healthcare, several studies have made significant contributions to the analysis of clinical notes using machine learning and deep learning techniques [48, 49, 50, 51]. In the field of cybersecurity, the importance of email filtering in protecting against spam and phishing threats cannot be overstated, leading numerous studies to explore the efficacy of deep learning approaches [16, 17, 18, 19, 52, 53]. In the sphere of business intelligence, scrutinizing customer feedback is paramount for grasping market dynamics and guiding strategic decisions, prompting a multitude of studies to delve into this area [20, 21, 22, 54, 55].

E. Arabic Text Classification

Text classification in Arabic has seen significant advancements with the application of machine learning and deep learning models. Recent studies have explored various aspects, from systematic reviews of methodologies and challenges specific to the Arabic language [56] to the implementation of deep learning models [57]. The complexity of Arabic text, with its rich morphology and syntax, poses unique challenges, addressed through semantic ontology-based approaches [58] and the adaptation of BERT models [59]. Moreover, the effectiveness of different machine learning algorithms in handling Arabic text classification tasks has been systematically reviewed, highlighting the need for more tailored approaches to better cater to the linguistic characteristics of Arabic [60].

3. CONCEPTS AND PRELIMINATIRES

This section serves as a glossary, equipping readers with the terminologies that are crucial for comprehending the nuanced aspects of few-shot learning methodologies.

A. Few-Shot Learning in Text Classification

Few-shot learning, especially in NLP, addresses learning from limited data. Wang et al. (2020) [25] provided a foundational survey of this concept in text classification, where labeled data scarcity is common.

Definition: Few-shot learning involves a program learning from a small set of experiences (E), tasks (T), and a performance metric (P), with its performance improving progressively on tasks T after exposure to experiences E. E is notably minimal in size [25].

B. Types of Few-Shot Learning Approaches

1) Prototypical Networks:

These networks simplify few-shot classification by creating a metric space based on distances to prototype representations of each class [27].

2) Matching Networks:

Designed to mimic human-like knowledge generalization, they adapt quickly to new tasks with minimal examples, utilizing memory-augmented neural networks [61, 28].

3) Relation Networks:

Operate on learning to compare small sets of images, discerning and generalizing from sparse data [62].

4) Meta-Learning:

Known as "learning to learn," models are trained to acquire foundational knowledge from various tasks and adapt rapidly with a limited set of labeled examples [62]. Model-Agnostic Meta-Learning (MAML) framework [64] introduced by Finn et al.in 2017, optimizes models for rapid adaptation.

5) Metric-Based Learning:

Learns a function to measure similarity or distance between data points, useful in semisupervised and unsupervised cross-domain FSL contexts. Notably, Vinyals et al. [64] introduce adaptive weighting for support set examples during one-shot query set classification, showcasing the effectiveness of metric-based techniques in Few-Shot Learning.

6) Transfer Learning:

Involves leveraging prior knowledge from related tasks to enhance performance in new, data-scarce tasks, effective when features are applicable to both tasks [65]. This approach is effective when features are applicable to both tasks, as noted by Yosinski et al. [66].

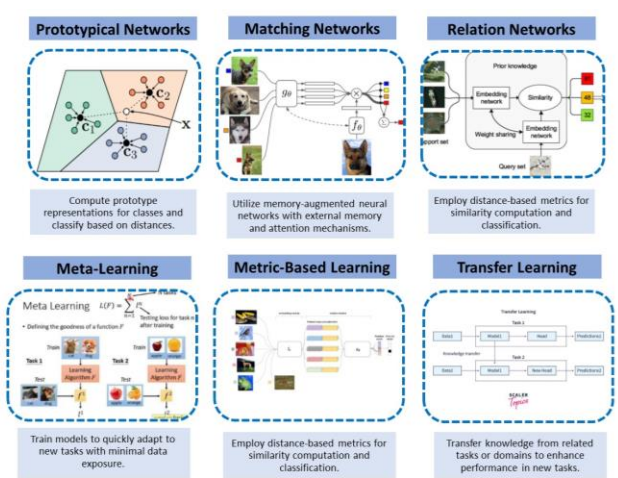

Figure 3. Types of few-shot learning approaches.

4. METHODOLOGY

In the next section, we will outline our methodology, focusing on the structured approach adopted for this research. We begin with an overview of our search strategy, highlighting the comprehensive and systematic methods used to identify relevant literature. This is followed by a discussion of the inclusion and exclusion criteria, explaining the principles guiding our selection of studies for detailed analysis. Lastly, we describe the study selection process, emphasizing the systematic steps and measures undertaken to ensure the reliability and validity of our approach.

A. Methodological Framework

In our experimental design, we have adopted a structured literature review approach utilizing the PRISMA guidelines to ensure a comprehensive and reproducible methodology [67]. Our PRISMA-based review commenced with an initial broad search, followed by screenings at various levels—title, abstract, and full text—to ascertain each study's relevance to our research question as shown in Figure 4. This meticulous process allowed us to filter the literature systematically, ensuring that the selected studies meet our inclusion criteria while maintaining a high standard of research integrity. The use of the PRISMA flow diagram will provide a visual representation of the search and selection process, detailing the number of records identified, included, and excluded, along with the reasons for exclusions. This robust experimental design will contribute significantly to the reliability and validity of our study findings. By focusing on recent literature, we aim to capture the latest advancements and trends and identify gaps for future research.

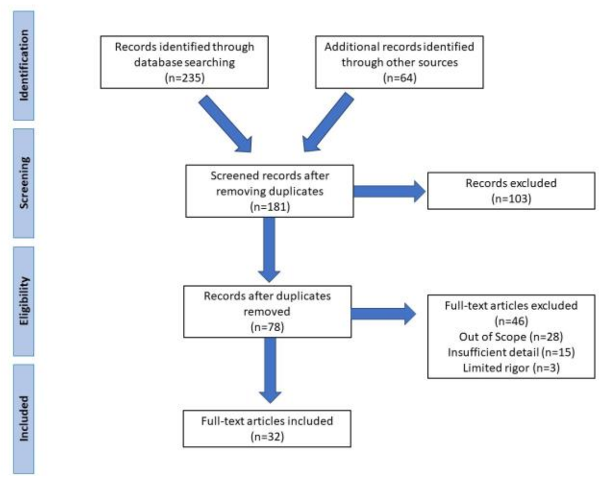

Figure 4. PRISMA Flow Diagram of the literature review process.

B. Databases Used for Literature Search

Our research leveraged a diverse array of academic databases to conduct a thorough and comprehensive search for relevant papers. We selected platforms renowned for their vast collections of scholarly articles, spanning diverse disciplines with a focus on AI and ML. These platforms are esteemed for rigorously curating and indexing peer-reviewed journals, conferences, and academic publications. By utilizing these databases, we ensured access to high-quality research for our literature review. Several electronic databases and search engines, including Google Scholar, IEEE Xplore, arXiv, ACM Digital Library, SpringerLink, ScienceDirect, PubMed, and Web of Science, were systematically queried for scholarly literature across various domains.

C. Search Strategy

This subsection details the systematic approach employed to identify relevant literature pertaining to FewShot Learning (FSL) in the realm of text classification. Our search protocol aimed to identify relevant and recent research, focusing on studies from the past six years to align with the latest developments. We targeted esteemed publishers' databases, prioritizing materials from respected conferences, journals, and workshops known for their contributions to AI and machine learning.

To refine our literature search strategy and ensure the identification of pertinent studies, we meticulously developed a keyword matrix based on a structured framework outlined in Table 1. This framework delineates key dimensions within the realm of natural language processing research, encompassing a range of application domains. For each of these domains, we compiled a targeted list of specific keywords, which were then thoughtfully combined with terms from the 'Technical Field', 'Learning Strategy', and 'Type of Learning' categories to enhance the specificity of our queries. For example, a search for papers in the health sector might leverage combinations such as "Patient care" with "text mining" and "Transfer Learning". This methodical approach to our search process allowed us to conduct a thorough and focused exploration of the relevant literature.

TABLE 1: OVERVIEW OF QUERY CATEGORIES AND CORRESPONDING KEYWORDS

Category	Queries
Technical Field	Natural language processing, text mining, text classification, named entity recognition, concept extraction
Learning Strategy	Zero-shot, one-shot, fewshot
Type of Learning	Meta-Learning, Transfer Learning, Metric-Based Learning, Matching networks, Prototypical network
Applications	
Sentiment Analysis	Opinion mining, Emotion Classification, Sentiment detection, Mood analysis
Health	Medical care, Health services, Patient care, Health systems
Business Intelligence	Decision support systems, Enterprise reporting, Business performance management, Data-driven decision-making
Spam Detection	Email spam filtering, Message classification, Unsolicited message blocking, Spam mail identification
Intent Detection	User intent identification, Query intention understanding, Purpose extraction, Intent-based response generation
Customer Feedback	Review analysis, Customer satisfaction measurement, Feedback sentiment assessment, Service improvement insights
Language Translation	Cross-lingual translation, Machine translation accuracy, Bilingual text alignment, Translation model optimization

D. Inclusion and Exlusion Criteria

Our research methodology was based on a precise set of inclusion criteria to ensure the selection of highly relevant studies. We reviewed peer-reviewed papers published in the last seven years, focusing on machine learning applications in areas like sentiment analysis, healthcare, business intelligence, and spam detection. We prioritized papers with empirical results for validation and included theoretical ones if they made significant contributions to the field. These criteria ensured the relevance and scholarly quality of our review.

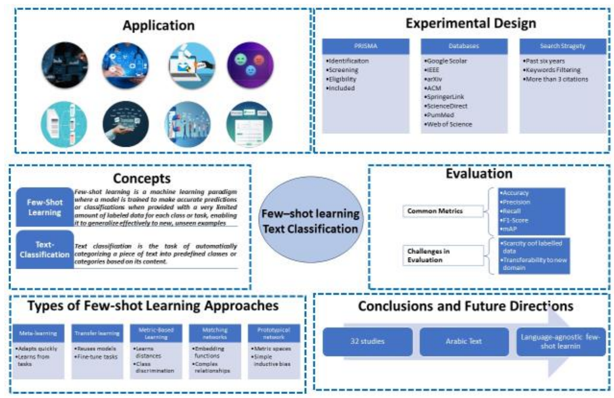

Figure 5. This figure illustrates the framework of our study, highlighting the key concepts and their interconnections pivotal to understanding our approach.

On the contrary, our exclusion criteria were meticulously crafted to sift through studies that did not align with our stringent standards for relevance and scientific robustness. Specifically, articles lacking peer review, such as pre-prints or editorial pieces, were disregarded. Furthermore, studies falling outside the purview of our defined technical domains or those published more than six years ago were omitted. Additionally, papers devoid of at least one citation were not excluded from consideration. Moreover, we excluded papers solely focused on text classification but not few-shot learning, as well as those addressing few-shot learning but not in the context of text classification. Lastly, articles unavailable in full-text format or inaccessible due to legal constraints were omitted from our review.

5. APPLICATIONS AND USE CASES

Few-shot learning addresses the challenge of limited labeled data in text classification, enabling classifiers to generalize effectively with minimal instances. It enhances sentiment analysis, conversational AI, and document categorization, improving customer experience, user interaction, and information management. Moreover, it provides robust solutions for spam detection, ensuring digital security and user trust.

TABLE 2: APPLICATIONS OF FEW-SHOT LEARNING (FSL) IN TEXT CLASSIFICATION

Application	Industry	Challenges
Sentiment Analysis	Marketing and Customer Service	Limited labeled sentiment datasets, domain adaptation
Topic Classification	News Aggregation Platforms	Generalizing across varied topics, adapting to trends
Spam Detection	Email Service Providers	Adapting to evolving spam tactics, limited labeled data
Intent Recognition	Chatbots and Virtual Assistants	Generalizing to diverse user intents, query variations
Named Entity Recognition	Information Extraction Systems	Generalizing to new entity types, handling ambiguity
Text Summarization	News Aggregation and Document Management	Capturing key information, coherent summaries
Document Classification	Legal and Regulatory Compliance	Adapting to legal jargon, generalizing across types
Fake News Detection	Social Media Platforms	Adapting to evolving misinformation tactics, multilingual

Table 2 elucidates the diverse applications of fewshot learning in text classification, encompassing a broad spectrum of industries. Complementing this, Figure 2 graphically represents the extensive reach and impact of few-shot learning across these varied applications, showcasing its transformative potential in enhancing text classification methodologies where labeled data is a luxury.

A. Benchmark Datasets

In the field of text classification, the development of deep learning algorithms has highlighted the necessity for varied datasets for effective model training. The surge in natural language processing (NLP) research has led to the creation of numerous datasets, both in English and Arabic, catering to the diverse needs of these advanced algorithms. These datasets are crucial to understanding the linguistic intricacies and training models for interpreting human language accurately. As researchers continue to compile and refine these collections, the aim is not only to advance text analysis techniques but also to ensure that models are versatile across different languages and cultural contexts, thereby enhancing the global applicability of NLP technologies.

Arabic Sentiment Analysis is an active research field, emphasizing the need for Arabic datasets. The motivation behind this endeavor stems from the realization that dominance of English datasets, have hindered the applicability of advanced ML and NLP models in Arabic contexts. Researchers are actively curating Arabic datasets, vital for training and evaluating models. Ongoing endeavors to expand Arabic datasets reflect a commitment to enhancing text classification and NLP in Arabic, addressing its linguistic and cultural intricacies. A summary of these datasets are presented the following table.

TABLE 3: OVERVIEW OF PUBLICLY AVAILABLE DATASET FOR TEXT CLASSIFICATION

Name	Year	Description	Size
Reuters-21578 [68]	1987	Categorized news articles	21,578 articles
Amazon Reviews [69]	2007	Product reviews and metadata from Amazon	34,686,770 reviews 6,643,669 users 2,441,053 products
PubMed dataset [70]	----	Biomedical literature	33 million records
IMDb Movie Reviews [71]	2011	Reviews for sentiment analysis	50,000 reviews
SpamAssassin Public Corpus [72]	2004	Labeled email messages	~6,000 messages
Sentiment140 [73]	2009	Tweets with sentiment annotations	1.6 million tweets
DBPedia Ontology Dataset [74]	2014	Wikipedia articles categorized	4.58 million articles
Yelp Reviews [75]	2015	Reviews for businesses and users	~4.7 million reviews
TREC Question Classification [76]	2002	Classified questions	~6,000 questions
Stanford Sentiment Treebank [77]	2013	Sentences annotated with sentiment	11,855 sentences
WikiQA [78]	2015	Q&A dataset for question answering systems	~3,000 QA pairs
Quora Question Pairs [79]	2017	Pairs of questions with annotations	~400,000 pairs
BBC News Dataset [80]	2006	Articles from BBC news website	2,225 articles
Jigsaw Toxic Comment Classification [81]	2018	Wikipedia comments with toxicity annotations	~160,000 comments
Multi-Domain Sentiment Dataset [69]	2007	Product reviews across types	~142,000 reviews
GameWikiSum [82]	2020	Dataset from online games for behavior analysis	~282,992 words
AG's News [83]	2015	Collection of news articles	~1 million
DBPedia [84]	2007	Structured content from Wikipedia	~6.0M entities
MIMIC II [85]	2016	Clinical data from ICU patients	12,000 clinical reports
Multigames dataset [86]	2017	Twitter data on game topics	12,780 tweets
Health Care Reform (HCR) [87]	2011	Tweets with sentiment labels	2,394 tweets
Sentiment Strength Twitter Dataset (SS-Tweet) [88]	2013	Tweets with sentiment strengths	4,242 tweets
SemEval Dataset	2013	Twitter sentiment analysis task	7,967 tweets
20News [89]	1995	News discussion forum discourse	18,000 posts
FewRel dataset [90]	2018	Relation classification dataset for FSL	100 relations, 70,000 instances
HuffPost [91]	2022	News headlines	210,000 headlines
EU legislation - EU-RLEX57K [92]	2019	EU legislative documents	57k documents
IMDB review [71]	2011	Movie reviews	50,000 reviews
Elec [93]	2013	Movie and product reviews	25,000 reviews

Name	Year	Description	Size
Yahoo! Answers [94]	2008	Q&A topic dataset	1.4 million
HUFF [91]	2022	News headlines	200,000 headlines
Banking77 [95]	2020	Intent classification dataset for banking domain	13,083 user utterances, 77 intents
HWU64 [96]	2019	Intent classification dataset across domains	25,478 sentences
Clinic150 [97]	2019	Intent classification dataset with out-of-scope queries	23,700 queries, including 22,500 in-scope
SST-2 [77]	2005	Sentiment labeled phrases in parse trees	215,154 phrases
GNAD [98]	2019	German news articles	10,000 articles
Head QA [99]	2019	Spanish health domain questions	not mentioned
Liu57 [96]	2019	User utterances from Amazon Mechanical Turk	25,478 utterances
Snippets [100]	2008	Web search snippets	Size not mentioned
Symptoms [42]	2019	Audio data of medical symptoms	8.5 hours of audio
NICTA PIBOSO [101]	2012	Benchmark for biomedical abstracts	11,616 sentences
NLPCC [143]	2017	Chinese news headlines dataset	Size not mentioned
THUCNews [143]	2011	News content and labels from Tsinghua University	Size not mentioned
AIStudio [91]	2022	News headline dataset for AIStudio competition	Over 240,000 news text
ArSarcasm-v2 [102]	2021	Arabic sarcasm detection and sentiment analysis	15,548 tweets
ArSAS [103]	2018	Arabic sentences with sentiment labels	21,000 sentences
ArTwitter [104]	2013	Arabic tweets with sentiment labels	1,975 tweets

B. Evaluation Metrics

Evaluation serves as a crucial aspect in assessing the performance of text classification models, encompassing both traditional text classification methods and emerging few-shot learning approaches. Common metrics such as precision (P), recall (R), F1-score, and accuracy play a pivotal role in quantifying the effectiveness of these models.

Precision (P): Measures the proportion of correctly predicted positive instances among all instances predicted as positive. It is calculated as

$$P = \frac{TP}{TP + FP}$$

Recall (R): Measures the proportion of correctly predicted positive instances among all actual positive instances. It is calculated as

$$P = \frac{TP}{TP + FN}$$

F1-score: Provides a balanced measure of a classifier's performance, calculated as

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

Accuracy: Measures the overall correctness of predictions made by the model across all classes. It is calculated as

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

Top-k Accuracy: Considers a prediction correct if the true class label is within the top-k predicted classes.

Mean Average Precision (mAP): Computes the average precision across all classes in multi-class classification tasks.

Harmonic Mean of Top-k Accuracy (HMTA): Combines the advantages of top-k accuracy and F1 score by computing the harmonic mean of the top-k accuracy and F1-score across all classes.

These metrics offer insights into the performance of text classification models, including few-shot learning approaches, enabling researchers to make informed decisions about their applicability in real-world scenarios. They aid in gauging effectiveness and robustness across domains. However, domain-specific metrics are vital for accurately capturing the intricacies of the task at hand.

6. Results

In this section, we present a comprehensive survey of our search and review papers pertaining to few-shot learning in text classification, as outlined in the subsequent table. Additionally, we conduct an analysis of the obtained results, comparing data collection outcomes and presenting relevant statistics.

A. Data Colleciton Results

In our study, we gathered a total of 75 research papers published after 2018, focusing on the domain of few-shot learning. However, upon applying our exclusion criteria, we narrowed down the selection to 32 papers. Notably, the majority of these exclusions, amounting to 30 papers, were related to studies conducted in English text contexts, while only two papers addressed issues in Arabic text. Out of the 30 studies conducted in English, one was conducted in German and Spanish, while an additional study was conducted in Chinese.

TABLE 4: PAPERS PERTAINING TO FEW-SHOT LEARNING IN TEXT CLASSIFICATION

Study	Year	Methodology	Dataset	Domain
Rios et al.[115]	2018	Novel approaches utilizing label space structure for few- and zero-shot learning.	MIMIC II, MIMIC III	Medical discharge summaries
Yan et al.[116]	2018	Siamese CNNs and few-shot learning-based framework for short text classification.	Multigames, HCR, SS-Tweet, SemEval-2013	Tweets about sentiment, games, and health care reform
Muhammad, Bailey and Chopra [117]	2018	Prototyping-based few-shot learning for classifying largely unlabeled document corpora.	20News	Web content from news
Yu et al. [118]	2018	Adaptive metric learning addressing challenges posed by task diversity in natural language tasks.	Amazon Reviews, User Intent Classification	Product reviews intent
Gao et al.[90]	2019	Introduction of FewRel dataset and its 100 Wikipedia-derived relations for few-shot relation classification.	FewRel	Wikipedia corpus and Wikidata knowledge bases
Ruiying Geng et al.[119]	2019	Induction Network leveraging dynamic routing in meta-learning for few-shot text classification.	Amazon Reviews	Product reviews, Open Domain Intent Classification
Bao et al. [120]	2019	Meta-learning incorporating distributional signatures for improved performance over prototypical networks in few-shot text classification.	20News, RCV, Reuters-2157, Amazon Reviews, HuffPost, FewRel	News, Reuters articles, customer reviews, headlines
Lu et al. [121]	2020	Multi-graph aggregation model using pre-trained word embeddings and predefined label relations for zero/few-shot document classification.	MIMIC II, MIMIC III, EU legislation dataset	Medical
Chalkidis et al.[122]	2020	Evaluation of LWANs, hierarchical methods, and Transformer-based approaches on diverse datasets addressing LMTC challenges.	EURLEX57K, MIMIC III, Amazon Reviews	English legislative documents, discharge summaries hospitals
Lwowski et al.[123]	2020	Utilization of few-shot learning to refine a semisupervised model with unlabeled COVID19 and labeled influenza datasets.	COVID-19	Tweets about COVID-19
Mukherjee and Awadallah [124]	2020	Enhanced self-training approach leveraging uncertainty estimates from Bayesian deep learning for semisupervised learning in text classification tasks.	IMDB, Dbpedia, AG News, Elec	Sentiment classification, product reviews, topic of news classification
Lyu et al.[125]	2020	Few-shot text classification approach combining pre-trained language models with an edgelabeling graph neural network for high performance.	Amazon Reviews, FewRel	Product reviews, relation
Lee et al.[126]	2021	Semi-supervised bootstrap learning enlarging training data using attention weights from an LSTM-based classifier and class-specific lexicons.	IMDB review, AG News, Yahoo, DBpedia	Sentiment analysis
Wei et al. [127]	2021	Data augmentation enhancing triplet networks and introducing a novel curriculum data augmentation strategy for accelerated training.	HUFF, FerWel, COV-C, Amazon Reviews	News headlines, relationship between specified head and tail tokens, COVID-related
Xia et al. [128]	2021	Incremental few-shot text classification with entailment approaches and benchmark datasets for incremental learning without re-training.	IFS-INTENT, IFS-RELATION	Intent detection, relation classification
Han et al. [129]	2021	Meta-learning framework combined with adversarial domain adaptation network for generating high-quality text embeddings in few-shot text classification.	Amazon Reviews, Reuters-21578	News, customer reviews, headlines
Lang et al. [130]	2021	TransPrompt, a transferable prompting framework for few-shot learning outperforming single-task and cross-task baselines across multiple datasets.	7 datasets	News, customer reviews, headlines
Min et al. [131]	2021	Noisy channel approach for language model prompting in few-shot text classification employing channel models for enhanced stability and performance.	11 datasets	Various scopes
Zhao et al. [132]	2022	MetaSLRCL, a meta-learning framework introducing dynamic learning rates and task-oriented curriculum learning for enhanced generalization.	FewRel, 20 News, DB Pedia Ontology	Relations, news articles, products
Chen et al. [133]	2022	ContrastNet, a contrastive learning framework enhancing class-specific text representations for improved few-shot text classification.	Banking77, HWU64, Clinic150, Liu57, HuffPost, Amazon Reviews, Reuters, 20News	News and reviews
Kim et al. [134]	2022	ALP method improving few-shot learning by generating augmented samples and introducing novel train-validation splitting strategies.	SST-2, IMDB, Elec, AG News, DBpedia	Sentiment classification, topic classification
Zhang et al. [135]	2022	PBML model integrating meta-learning and prompt-tuning for tackling challenges in few-shot text classification, achieving state-of-the-art performance.	FewRel, HuffPost, Reuters, Amazon Reviews	Relations, news articles, products

Study	Year	Methodology	Dataset	Domain
Liu et al. [136]	2022	Meta-FCS, a few-shot short-text classification method efficiently transferring common features across diverse fields while emphasizing task-specific features.	Amazon Reviews, FewRel	Product reviews, relation
Muller et al. [137]	2022	FASL, a platform integrating few-shot learning and active learning techniques for accelerated model training	AG News, GNAD, Head QA, Amazon Reviews	News in English and German, Spanish catalogue product reviews
Kim [138]	2022	LST, a self-training method leveraging a lexicon to guide the pseudolabeling process and improve generalization.	SST-2, IMDB, Elec, AG News, Dbpedia	Sentiment classification
Yu et al. [139]	2023	Enhancement of retrieval-augmented methods for few-shot learning by developing task-specific retrieval metrics and introducing EM-L and R-L training objectives.	10 diverse datasets	Sentiment classification, Natural Language Inference, aspect-based sentiment analysis
Li [140]	2023	A novel approach incorporating knowledge distillation and graph aggregation for improved performance in few-shot text classification.	HuffPost, 20News, Reuters	News
Liu et al.[141]	2023	Novel strategies for few-shot text classification distribution estimation using unlabeled query samples and Gaussian assumption for class/sample distributions.	HuffPost, Amazon Reviews, Reuter, 20News, Banking77, Liu57, Clinic150	News, product reviews, medical
Liao [142]	2023	Mask-BERT enhances few-shot learning capabilities by selectively applying masks on text inputs to guide attention towards discriminative tokens.	AG News, Dbpedia14, Snippets, Symptoms, PubMed20k, NICTA-PIBOSO	Various domains
Wang [143]	2023	Knowledge-guided prompt learning approach leveraging implicit knowledge in pre-trained language models for improved performance.	NLPCC, THUCNews	English and Chinese news
Seham Basabain et. al [144]	2023	Transformer-based models like AraBERT, particularly effective for tasks with limited data	ArSarcasm-v2, ArSAS, ArTwitter, Ans, SAtour	Arabic news, tweets, and sarcasm context
Muhammad Khalifa et al [145]	2021	Self-training approach utilizing pre-trained language models in zero- and few-shot scenarios to improve performance on data-scarce language varieties.	Arabic named entity recognition (NER) and part-of-speech (POS) tagging	-

This observation underscores the predominant focus on English language studies within the few-shot learning literature, highlighting a potential gap in research attention towards other linguistic contexts, particularly Arabic. The frequency distribution of the number of papers over time from our study is depicted in Figure 6.

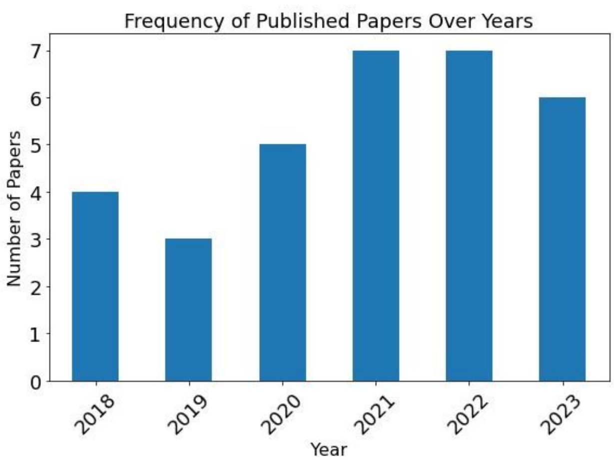

Figure 6. Number of papers related to few-shot learning in text classification from 2018 to 2013.

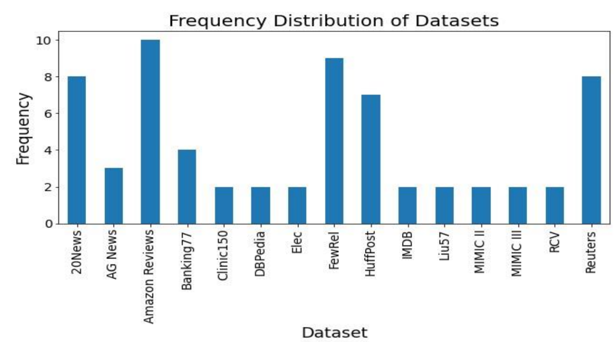

Figure 7. The frequency distribution of publicly available datasets for text classification and few-shot learning utilized in our surveyed papers.

1) Summary of datasets

We undertook a comprehensive analysis of the datasets employed across the spectrum of research papers addressing text classification and few-shot learning. The figure selectively highlights datasets that were referenced more than once in the surveyed literature, deliberately excluding those with only a single mention to focus on the most recurrently used resources in the field.

Exploring through the datasets, the 'Amazon Reviews' dataset stands out with ten citations, owing to its diverse user-generated content, making it invaluable for text classification studies. '20News' and 'FewRel' datasets are prominent with eight mentions each, indicating their importance in both text and relation classification tasks, especially in few-shot learning scenarios. 'HuffPost' and

'Reuters' datasets follow in frequency, with seven and five mentions respectively.

B. Summary of methodologies and applications

Surveyed methodologies converge on tackling labeled data scarcity in NLP, leveraging deep learning. Many studies opt for neural network architectures tailored for few-shot learning. Meta-learning is prominent, stressing adaptable models with minimal training data. Innovations include metric learning to optimize similarity measures and enhance few-shot system performance. Others explore dynamic routing within meta-learning frameworks to craft class-agnostic representations, bolstering few-shot classification efficacy.

Graph neural networks are employed to capture the relational structure inherent in text data, which is particularly beneficial in tasks like relation extraction and document classification. Selftraining and semi-supervised learning methods are also prominent, leveraging the abundance of unlabeled data to enhance model performance. Additionally, contrastive learning has been utilized to refine text representations, enabling models to distinguish between classes more effectively in a few-shot setting

C. Innovation and Advancements in Few-Shot Learning and Text classifications

Innovation in methodology is a standout feature of the surveyed literature. Several studies have proposed adaptive models that tailor their approach based on the specifics of the task at hand, whether through the use of hierarchical methods that account for label dependencies or through models that incorporate attention mechanisms to focus on the most salient features of the text. The use of pre-trained embeddings is a common strategy to bootstrap the learning process, leveraging the rich representations captured by models like BERT, which have been pre-trained on vast corpuses of text.

Methods like prototypical networks and meta-learning demonstrate efforts to develop models learning abstract representations from few examples and applying them to new tasks or classes, indicating a commitment to versatile learning paradigms.

D. Arabic Text Classification

The exploration of Arabic text classification within the surveyed papers, although less prevalent, signifies an emergent interest in applying few-shot learning principles to less-resourced languages. The studies that do focus on Arabic tackle the unique challenges of dialectal variation and the lack of extensive labeled datasets by deploying Transformer-based models and self-training methods. While the research on Arabic represents only a very small fraction of the overall body of work, it is a critical step toward achieving linguistic diversity in NLP and addressing the research gap in non-English languages.

7. Conclusions

Our comprehensive review of the literature in few-shot learning within the realm of text classification has identified a total of 32 studies that meet our stringent criteria, all conducted post-2018. Few-shot learning methodologies demonstrate significant potential in leveraging minimal labeled data for effective model training. They mitigate constraints of extensive annotated datasets, enhancing adaptability and efficiency in text classification across diverse domains. By enabling models to learn from fewer examples, they reduce reliance on large-scale data and associated acquisition costs. These findings highlight few-shot learning's capability to diminish dependency on voluminous annotated datasets, traditionally essential for training robust text classification models, thus showcasing its transformative impact on model development and resource utilization.

Despite the evident progress and achievements, our review delineates notable gaps within the extant research on few-shot learning for text classification. A conspicuous emphasis on English-centric studies emerges, with a stark underrepresentation of other languages, particularly those with intricate linguistic frameworks such as Arabic. This skewed focus inadvertently sidelines the unique challenges and opportunities inherent in non-English languages, which could otherwise contribute valuable perspectives to the adaptability and efficacy of few-shot learning paradigms. Moreover, the proliferation of datasets and benchmarks catering to text classification and broader NLP tasks has not been paralleled by the development of resources specifically designed for few-shot learning contexts, especially within multilingual settings. This discrepancy highlights an urgent need for creation of datasets and the establishment of evaluation benchmarks, aimed at fostering a deeper and more nuanced understanding of few-shot learning's capabilities and limitations across varied linguistic landscapes.

In light of the identified gaps, we propose several avenues for future research aimed at advancing the field of few-shot learning in text classification. A critical imperative is the diversification of few-shot learning research to encompass a broader spectrum of languages, with a special focus on languages such as Arabic that have been historically marginalized in NLP research. This endeavor could involve the formulation of novel datasets and benchmarks that are specifically tailored to accommodate the linguistic diversity and complexity of various languages. Further exploration into the realm of cross-lingual and language-agnostic fewshot learning models holds the promise of developing universally applicable and scalable solutions for text classification. Additionally, the convergence of fewshot learning with cutting-edge technologies, including transformer-based architectures and meta-learning strategies, presents fertile ground for enhancing the generalizability, efficiency, and overall performance of text classification models.

The ramifications of our literature survey stretches beyond academia, shaping future research and NLP deployment. Few-shot learning's growth in text classification can democratize access to advanced NLP, benefiting resource-limited environments and languages. This inclusivity promotes balanced technological innovation across cultures. Sustained research efforts are vital for addressing current limitations and exploring new possibilities in text classification and NLP's future landscape.

In essence, the exploration of few-shot learning within text classification has led to important discoveries and greatly benefited the wider field of natural language processing (NLP). Tackling the current shortcomings and exploring new areas could lead to significant breakthroughs, making NLP tools more available, effective, and inclusive for a diverse range of languages.

8. REFERENCES

[1] C. D. Manning, M. Surdeanu, et al., The stanford corenlp natural language processing toolkit (2014) 55–60.

[2] Y. LeCun, et al., Deep learning, Nature 521 (2015) 436–444.

[3] T. Young, D. Hazarika, S. Poria, E. Cambria, Recent trends in deep learning based natural language processing, IEEE Computational Intelligence Magazine 13 (2018) 55–75.

[4] C. C. Aggarwal, C. Zhai, A survey of text classification algorithms (2012) 163–222.

[5] A. Younus, et al., Context-aware hybrid classification system for fine-grained arabic text categorization, Journal of Information Science 40 (2014) 540–554.

[6] F. Sebastiani, Machine learning in automated text categorization, ACM computing surveys (CSUR) 34 (2002) 1–47.

[7] E. Agichtein, E. Brill, S. Dumais, R. Ragno, Improving web search ranking by incorporating user behavior information, in: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 2006, pp. 19–26.

[8] V. Kumar, S. Kumar, Automated customer service agent, in: Proceedings of the International Conference on Data Engineering and Communication Technology, volume 1, Springer, 2016, pp. 655–661.

[9] X. Liu, T. Shi, G. Zhou, et al., Emotion classification for short texts: an improved multi-label method, Humanities and Social Sciences Communications 10 (2023) 306. URL: https: //www.nature.com/articles/s41599-023-01816-6. doi:10.1057/s41599-023-01816-6.

[10] A. Alqahtani, H. Ullah Khan, S. Alsubai, M. Sha, A. Almadhor, T. Iqbal, S. Abbas, An efficient approach for textual data classification using deep learning, Frontiers in Computational Neuroscience 16 (2022) 992296.

[11] M. H. Abd El-Jawad, R. Hodhod, Y. M. Omar, Sentiment analysis of social media networks using machine learning, in: 2018 14th international computer engineering conference (ICENCO), IEEE, 2018, pp. 174–176.

[12] C. Dhaoui, C. M. Webster, L. P. Tan, Social media sentiment analysis: lexicon versus machine learning, Journal of Consumer Marketing 34 (2017) 480–488.

[13] I. Spasic, G. Nenadic, Clinical text data in machine learning: Systematic review, JMIR Medical Informatics 8 (2020) e17984. URL: https://pubmed.ncbi.nlm.nih.gov/32229465/. doi:10.2196/17984.

[14] S. Hussain, S. Lee, Semantic transformation model for clinical documents in big data to support healthcare analytics, in: 2015 Tenth International Conference on Digital Information Management (ICDIM), IEEE, 2015, pp. 99–102.

[15] A. Kormilitzin, et al., Med7: A transferable clinical natural language processing model for electronic health records, Artificial Intelligence in Medicine 118 (2021) 102086.

[16] G. V. Cormack, T. R. Lynam, Online supervised spam filter evaluation, ACM Transactions on Information Systems (TOIS) 25 (2007) 11.

[17] M. Sokolov, Applied Machine Learning for Cybersecurity in Spam Filtering and Malware Detection, East Carolina University, 2020.

[18] S. Magdy, Y. Abouelseoud, M. Mikhail, Efficient spam and phishing emails filtering based on deep learning, Computer Networks 206 (2022) 108826.

[19] T. Gangavarapu, C. Jaidhar, B. Chanduka, Applicability of machine learning in spam and phishing email filtering: review and approaches, Artificial Intelligence Review 53 (2020) 5019– 5081.

[20] B. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (2008) 1–135.

[21] J. P. Bharadiya, Machine learning and ai in business intelligence: Trends and opportunities, International Journal of Computer (IJC) 48 (2023) 123–134.

[22] W. A. Khan, et al., Machine learning facilitated business intelligence (part i) neural networks learning algorithms and applications, Industrial Management & Data Systems 120 (2020) 164–195.

[23] C. Sun, A. Shrivastava, S. Singh, A. Gupta, Revisiting unreasonable effectiveness of data in deep learning era, Proceedings of the IEEE international conference on computer vision (2017) 843–852.

[24] D. Amodei, et al., Deep speech 2: End-to-end speech recognition in english and mandarin, International conference on machine learning (2016) 173–182.

[25] Y. Wang, et al., Generalizing from a few examples: A survey on few-shot learning, ACM computing surveys (csur) 53 (2020) 1–34.

[26] L. Fei-Fei, R. Fergus, P. Perona, One-shot learning of object categories, IEEE transactions on pattern analysis and machine intelligence 28 (2006) 594–611.

[27] J. Snell, K. Swersky, R. Zemel, Prototypical networks for fewshot learning, Advances in neural information processing systems 30 (2017).

[28] O. Vinyals, et al., Matching networks for one shot learning, Advances in neural information processing systems 29 (2016) 3630–3638.

[29] Y. Ge, et al., Fewshot learning for medical text: A systematic review, Preprint on ResearchGate (2022). URL: https://www.researchgate. net/publication/360261240_Few-shot_learning_for_medical_text_A_systematic_review.

[30] A. Le Glaz, et al., Machine learning and natural language processing in mental health: systematic review, Journal of Medical Internet Research 23 (2021) e15708.

[31] J. Yang, et al., A survey of knowledge enhanced pre-trained models, arXiv preprint arXiv:2110.00269 (2021).

[32] S. Minaee, et al., Deep learning–based text classification: a comprehensive review, ACM computing surveys (CSUR) 54 (2021) 1–40.

[33] Q. Li, et al., A survey on text classification: From shallow to deep learning, arXiv preprint arXiv:2008.00364 (2020).

[34] Q. Li, H. Peng, et al., A survey on text classification: From traditional to deep learning, ACM Transactions on Intelligent Systems and Technology (TIST) 13 (2022) 1–41.

[35] H. Wu, Y. Liu, J. Wang, Review of text classification methods on deep learning., Computers, Materials & Continua 63 (2020).

[36] G. Brauwers, F. Frasincar, A survey on aspect-based sentiment classification, ACM Computing Surveys 55 (2022) 1–37.

[37] A. Mabrouk, et al., Deep learningbased sentiment classification: A comparative survey, IEEE Access 8 (2020) 85616–85638.

[38] Q. Ye, B. Y. Lin, X. Ren, Crossfit: A few-shot learning challenge for cross-task generalization in nlp, arXiv preprint arXiv:2104.08835 (2021).

[39] J. Bragg, et al., Flex: Unifying evaluation for few-shot nlp, Advances in Neural Information Processing Systems 34 (2021) 15787–15800.

[40] Y. Hou, et al., Fewjoint: A few-shot learning benchmark for joint language understanding, arXiv preprint arXiv:2009.08138 (2020).

[41] E. Perez, D. Kiela, K. Cho, True few-shot learning with language models, Advances in neural information processing systems 34 (2021) 11054–11070.

[42] X. V. Lin, et al., Few-shot learning with multilingual language models, arXiv preprint arXiv:2112.10668 (2021).

[43] S. Soltan, et al., Alexatm 20b: Few-shot learning using a large-scale multilingual seq2seq model, arXiv preprint arXiv:2208.01448 (2022).

[44] E. M. B. Nagoudi, et al., Jasmine: Arabic gpt models for few-shot learning, arXiv preprint arXiv:2212.10755 (2022).

[45] M. U. Salur, I. Aydin, A novel hybrid deep learning model for sentiment classification, IEEE Access 8 (2020) 58080–58093.

[46] Z. Hameed, B. Garcia-Zapirain, Sentiment classification using a single-layered bilstm model, Ieee Access 8 (2020) 73992–74001.

[47] S. Tam, R. B. Said and Ö. Ö. Tanriöver, "A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification," in IEEE Access, vol. 9, pp. 41283-41293, 2021, doi: 10.1109/ACCESS.2021.3064830

[48] A. Mustafa, M. Rahimi Azghadi, Automated machine learning for healthcare and clinical notes analysis, Computers 10 (2021) 24.

[49] V. Kumar, et al., Ensembling classical machine learning and deep learning approaches for morbidity identification from clinical notes, IEEE Access 9 (2020) 7107–7126.

[50] H. Lu, L. Ehwerhemuepha, C. Rakovski, A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance, BMC medical research methodology 22 (2022) 181.

[51] Q. Wei, et al., A study of deep learning approaches for medication and adverse drug event extraction from clinical text, Journal of the American Medical Informatics Association 27 (2020) 13–21.

[52] P. Verma, A. Goyal, Y. Gigras, Email phishing: Text classification using natural language processing, Computer Science and Information Technologies 1 (2020) 1–12.

[53] F. Janez-Martino, et al., Classification of spam emails through hierarchical clustering and supervised learning, arXiv preprint arXiv:2005.08773 (2020).

[54] A. Kumar, V. Dabas, P. Hooda, Text classification algorithms for mining unstructured data: a swot analysis, International Journal of Information Technology 12 (2020) 1159–1169.

[55] Z. Desai, K. Anklesaria, H. Balasubramaniam, Business intelligence visualization using deep learning based sentiment analysis on amazon review data, in: 2021 12th International Conference on computing communication and Networking Technologies (ICCCNT), IEEE, 2021, pp. 1–7.

[56] Abiodun, et al., (2021). A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities. Neural Computing and Applications, 33(22), 15091-15118.

[57] Alwehaibi A, et al., A study of the performance of embedding methods for Arabic short-text sentiment analysis using deep learning approaches. Journal of King Saud University-Computer and Information Sciences. 2022 Sep 1;34(8):6140-9.

[58] Yuvaraj D, et al.. Ontology Based Semantic Enrichment for Improved Information Retrieval Model. International Journal of Intelligent Systems and Applications in Engineering. 2024 Feb 7;12(15s):70-7.

[59] Alshahrani N, et al.. Arabic Synonym BERT-based Adversarial Examples for Text Classification. arXiv preprint arXiv:2402.03477. 2024 Feb 5.

[60] Al-Qerem A, et al., Utilizing Deep Learning Models (RNN, LSTM, CNN-LSTM, and Bi-LSTM) for Arabic Text Classification. InArtificial Intelligence-Augmented Digital Twins: Transforming Industrial Operations for Innovation and Sustainability 2024 Jan 20 (pp. 287-301). Cham: Springer Nature Switzerland.

[61] J. Weston, S. Chopra, A. Bordes, Memory networks, arXiv preprint arXiv:1410.3916 (2014).

[62] F. Sung, et al., Learning to compare: Relation network for fewshot learning, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1199–1208.

[63] T. Hospedales, et al., Metalearning in neural networks: A survey, IEEE transactions on pattern analysis and machine intelligence 44 (2021) 5149– 5169.

[64] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: International conference on machine learning, PMLR, 2017, pp. 1126–1135.

[65] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on knowledge and data engineering 22 (2009) 1345–1359.

[66] J. Yosinski, et al., How transferable are features in deep neural networks?, Advances in neural information processing systems 27 (2014).

[67] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, P. Group*, Preferred reporting items for systematic reviews and metaanalyses: the prisma statement, Annals of internal medicine 151 (2009) 264–269.

[68] P. J. Hayes, S. P. Weinstein, Construe/tis: A system for contentbased indexing of a database of news stories., in: IAAI, volume 90, 1990, pp. 49–64.

[69] J. Blitzer, et al., Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification, in: Proceedings of the 45th annual meeting of the association of computational linguistics, 2007, pp. 440–447.

[70] PubMed Database, https://pubmed.ncbi.nlm.nih.gov/, Accessed: 2024-01-31.

[71] A. Maas, et al., Learning word vectors for sentiment analysis, in: Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies, 2011, pp. 142–150.

[72] A. Go, et al., Sentiment140: A Twitter Sentiment Analysis Tool, 2009. Available: http://www. sentiment140.com/.

[73] DBpedia Ontology Dataset, http://dbpedia.org/ ontology/, Accessed: 2024-01-31.

[74] N. Asghar, Yelp dataset challenge: Review rating prediction, arXiv preprint arXiv:1605.05362 (2016).

[75] X. Li, D. Roth, Learning question classifiers, in: COLING 2002: The 19th International Conference on Computational Linguistics, 2002.

[76] R. Socher, et al., Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2013. URL: https://nlp.stanford.edu/sentiment/.

[77] Y. Yang, et al., Wikiqa: A challenge dataset for open-domain question answering, in: Proceedings of the 2015 conference on empirical methods in natural language processing, 2015, pp. 2013–2018.

[78] Quora question pairs, https://data.quora.com/First-Quora-Dataset-Release-Question-Pairs, 2017. Accessed: 2024-01-31.

[79] D. Greene, P. Cunningham, BBC News Dataset, http://mlg. ucd.ie/datasets/bbc.html, 2006. Accessed: 2024-01-31. [81] Jigsaw Toxic Comment Classification Dataset, https://www.kaggle.com/c/ jigsaw-toxic-comment-classification-challenge, 2018. Accessed: 2024-01-31.

[80] D. Antognini, B. Faltings, Gamewikisum: a novel large multi-document summarization dataset, arXiv preprint arXiv:2002.06851 (2020).

[81] X. Zhang, J. Zhao, Y. LeCun, Character-level convolutional networks for text classification, Advances in neural information processing systems 28 (2015).

[82] S. Auer, et al., Dbpedia: A nucleus for a web of open data, in: international semantic web conference, Springer, 2007, pp. 722–735.

[83] A. E. Johnson, et al., Mimic-iii, a freely accessible critical care database, Scientific data 3 (2016) 1–9.

[84] L. Yan, et al., Learning discriminative sentiment chunk vectors for twitter sentiment analysis, 18 (2017) 1605–1613.

[85] M. Speriosu, et al., Twitter polarity classification with label propagation over lexical links and the follower graph, in: Proceedings of the First workshop on Unsupervised Learning in NLP, 2011, pp. 53–63.

[86] H. Saif, et al., Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the sts-gold (2013).

[87] K. Lang, Newsweeder: Learning to filter netnews, in: Machine learning proceedings 1995, Elsevier, 1995, pp. 331–339.

[88] X. Han, H. Zhu, P. Yu, Z. Wang, Y. Yao, Z. Liu, M. Sun, Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation, arXiv preprint arXiv:1810.10147 (2018).

[89] R. Misra, News category dataset, arXiv preprint arXiv:2209.11429 (2022).

[90] I. Chalkidis, et al., Large-scale multi-label text classification on eu legislation, arXiv preprint arXiv:1906.02192 (2019).

[91] J. McAuley, J. Leskovec, Hidden factors and hidden topics: understanding rating dimensions with review text, in: Proceedings of the 7th ACM conference on Recommender systems, 2013, pp. 165–172.

[92] M. W. Chang, et al., Importance of semantic representation: Dataless classification., in: Aaai, volume 2, 2008, pp. 830–835.

[93] I. Casanueva, T. Temcinas, D. Gerz, M. Henderson, I. Vuliˇ c,´ Efficient intent detection with dual sentence encoders, arXiv preprint arXiv:2003.04807 (2020).

[94] X. Liu, et al., Benchmarking natural language understanding services for building conversational agents, arXiv preprint arXiv:1903.05566 (2019).

[95] S. Larson, et al., An evaluation dataset for intent classification and out-of-scope prediction, arXiv preprint arXiv:1909.02027 (2019).

[96] T. Block, M. Domer, Ten thousand german news articles¨ dataset, 2019.

[97] D. Vilares, C. Gomez-Rodr´ ´ıguez, Head-qa: A healthcare dataset for complex reasoning, arXiv preprint arXiv:1906.04701 (2019).

[98] X. H. Phan, et al., Learning to classify short and sparse text & web with hidden topics from largescale data collections, in: Proceedings of the 17th international conference on World Wide Web, 2008, pp. 91–100.

[99] I. Amini, et al., Overview of the alta 2012 shared task, in: Proceedings of the Australasian Language Technology Association Workshop 2012, 2012, pp. 124–129.

[100] I. A. Farha, et al., Overview of the wanlp 2021 shared task on sarcasm and sentiment detection in arabic, in: Proceedings of the sixth Arabic natural language processing workshop, 2021, pp. 296–305.

[101] A. Elmadany, et al., Arsas: An arabic speech-act and sentiment corpus of tweets, OSACT 3 (2018) 20.

[102] T. Hailat, et al., E. E. T. A. M. Translators, Ieee jordan conference on applied electrical engineering and computing technologies (aeect 2013)-it applications & systems, Amman, Jordan (2013).

[103] J. Khouja, Stance prediction and claim verification: An Arabic perspective, arXiv preprint arXiv:2005.10410 (2020).

[104] M. Nabil, M. Aly, A. Atiya, Astd: Arabic sentiment tweets dataset, in: Proceedings of the 2015 conference on empirical methods in natural language processing, 2015, pp. 2515–2519.

[105] A. Mohammed, R. Kora, Deep learning approaches for arabic sentiment analysis, Social Network Analysis and Mining 9 (2019) 1–12.

[106] K. M. Alomari, et al., Arabic tweets sentimental analysis using machine learning, in: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Springer, 2017, pp. 602–610.

[107] M. Aly, A. Atiya, Labr: A large scale arabic book reviews dataset, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2013, pp. 494–498.

[108] I. Zeroual, et al., Osian: Open source international arabic news corpus-preparation and integration into the clarin-infrastructure, in: Proceedings of the fourth arabic natural language processing workshop, 2019, pp. 175–182.

[109] Y. Benajiba, M. Diab, Arabic named entity recognition corpus, in: Proceedings of the Fourth Linguistic Annotation Workshop, Association for Computational Linguistics, 2007, pp. 37–40.

[110] T. Mitchell, et al., Ace 2003: Automatic content extraction evaluation, in: Proceedings of the 6th Conference on Language Resources and Evaluation (LREC 2008), European Language Resources Association (ELRA), 2003, pp. 11–17.

[111] B. Alharbi, et al., Asad: A twitterbased benchmark arabic sentiment analysis dataset, arXiv preprint arXiv:2011.00578 (2020).

[112] S. A. Chowdhury, et al., Improving arabic text categorization using transformer training diversification, in: Proceedings of the Fifth Arabic Natural Language Processing Workshop, 2020, pp. 226–236.

[113] A. Rios, R. Kavuluru, Few-shot and zero-shot multi-label learning for structured label spaces, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing, volume 2018, NIH Public Access, 2018, p. 3132.

[114] L. Yan, et al., Few-shot learning for short text classification, Multimedia Tools and Applications 77 (2018) 29799–29810.

[115] K. Bailey, et al., Few-shot text classification with pretrained word embeddings and a human in the loop, arXiv preprint arXiv:1804.02063 (2018).

[116] M. Yu, et al., Diverse few-shot text classification with multiple metrics, arXiv preprint arXiv:1805.07513 (2018).

[117] R. Geng, B. Li, et al., Induction networks for few-shot text classification, arXiv preprint arXiv:1902.10482 (2019).

[118] Y. Bao, M. Wu, et al., Few-shot text classification with distributional signatures, arXiv preprint arXiv:1908.06039 (2019).

[119] J. Lu, L. Du, et al., Multi-label few/zeroshot learning with knowledge aggregated from multiple label graphs, arXiv preprint arXiv:2010.07459 (2020).

[120] I. Chalkidis, et al., An empirical study on largescale multi-label text classification including few and zero-shot labels, arXiv preprint arXiv:2010.01653 (2020).

[121]B. Lwowski, P. Najafirad, Covid-19 surveillance through twitter using self-supervised and few shot learning, in: Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020, 2020.

[122]S. Mukherjee, A. Awadallah, Uncertainty-aware self-training for few-shot text classification, Advances in Neural Information Processing Systems 33 (2020) 21199–21212.

[123]C. Lyu, et al., Few-shot text classification with edge-labeling graph neural network-based prototypical network, in: Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 5547–5552.

[124]J. H. Lee, et al., Salnet: Semi-supervised fewshot text classification with attention-based lexicon construction, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, 2021, pp. 13189–13197.

[125]J. Wei, et al., Few-shot text classification with triplet networks, data augmentation, and curriculum learning, arXiv preprint arXiv:2103.07552 (2021).

[126]C. Xia, et al., Incremental few-shot text classification with multi-round new classes: Formulation, dataset and system, arXiv preprint arXiv:2104.11882 (2021).

[127]C. Han, et al., Metalearning adversarial domain adaptation network for few-shot text classification, arXiv preprint arXiv:2107.12262 (2021).

[128]C. Wang, et al., Transprompt: Towards an automatic transferable prompting framework for few-shot text classification, in: Proceedings of the 2021 conference on empirical methods in natural language processing, 2021, pp. 2792–2802.

[129]S. Min, M. Lewis, et al., Noisy channel language model prompting for few-shot text classification, arXiv preprint arXiv:2108.04106 (2021).

[130]K. Zhao, X. Jin, et al., Metaslrcl: A selfadaptive learning rate and curriculum learning based framework for few-shot text classification, in: Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 2065–2074.

[131]J. Chen, R. Zhang, et al., Contrastnet: A contrastive learning framework for few-shot text classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, 2022, pp. 10492–10500.

[132]H. H. Kim, D. Woo, et al., Alp: Data augmentation using lexicalized pcfgs for few-shot text classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, 2022, pp. 10894–10902.

[133]H. Zhang, et al., Prompt-based metalearning for few-shot text classification, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 1342–1357.

[134]W. Liu, J. et al., Few-shot short-text classification with language representations and centroid similarity, Applied Intelligence 53 (2023) 8061–8072.

[135]T. Muller, G. P¨ erez-Torr´ o, A. Basile, M. Franco-Salvador, Active few-shot learning with fasl, in: International Conference on Applications of Natural Language to Information Systems, Springer, 2022, pp. 98–110.

[136]H. Kim, et al., Lst: Lexicon-guided selftraining for few-shot text classification, arXiv preprint arXiv:2202.02566 (2022).

[137]G. Yu, L. Liu, et al., Retrieval-augmented few-shot text classification, in: Findings of the Association for Computational Linguistics: EMNLP 2023, 2023, pp. 6721–6735.

[138]H. Li, G. Huang, et al., Seml: Self-supervised information-enhanced meta-learning for few-shot text classification, International Journal of Computational Intelligence Systems 16 (2023) 111.

[139]H. Liu, F. Zhang, et al., Boosting few-shot text classification via distribution estimation, arXiv preprint arXiv:2303.16764 (2023).

[140]W. Liao, Z. Liu, et al., Mask-guided bert for few shot text classification, arXiv preprint arXiv:2302.10447 (2023).

[141]L. Wang, R. Chen, L. Li, Knowledge-guided prompt learning for few-shot text classification, Electronics 12 (2023) 1486.

[142]S. Basabain, et al., Enhancing arabic-text feature extraction utilizing label-semantic augmentation in few/zero-shot learning, Expert Systems (2023) e13329.

[143]M. Khalifa, M. Abdul-Mageed, et al., Self-training pretrained language models for zero-and few-shot multi-dialectal arabic sequence labeling, arXiv preprint arXiv:2101.04758 (2021).

Amani Aljehani, a research scholar at King Abdulaziz University, specializes in machine learning and artificial intelligence. With a focus on Information Systems, she seeks innovative solutions to real-world challenges in the field of computer science.

Syed Hamid Hasan is a Professor of Computer Science at the Faculty of Computing & Information Technology (FCIT) of King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia. The area of his interests in the research work are Data Science, E-Security, E-Business, E-Learning, TQM. He has also contributed two book chapters. He has a number of publications to his credit in WOS journals.

Usman Ali Khan is an Associate Professor in the Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. He received his Ph.D. in Software Engineering from the Integral University, India.
His current research interest includes AI, Machine Learning, Software Engineering and Project Management He has published many research papers at national and international forums.