Annals of Emerging Technologies in Computing (AETiC)
Vol. 8, No. 4, 2024

Review Article

A Systematic Literature Review on Cross Domain Sentiment Analysis Techniques: PRISMA Approach

Rati Sharma* **and Kamlesh Lakhwani**

JECRC University, Jaipur, India
Rati.sharma1510@gmail.com; kamlesh.lakhwani@jecrcu.edu.in
*Correspondence: Rati.sharma1510@gmail.com

Received: 22nd May 2023; Accepted: 29th August 2024; Published: 1st October 2024

Abstract: Cross Domain Sentiment Analysis (CDSA) is a method that uses rich and quality-labeled source domain data to identify the sentiments of poorly or without labeled target data. In the past decade, ample research studies have focused on this topic to solve and propose efficient CDSA methods. However, an extensive investigation of these past studies is required to find a window of improvement. The main aim of the study is to figure out considerable developments, methodologies, and SOTA techniques in the recent past. This research study presents a systematic literature review to analyze the CDSA studies published from 2017 to 2023. The authors have selected 34 articles overall and categorized them into seven different CSDA techniques. The extensive analysis of these studies' results (in the form of graphs and tables) into different parameters that impact the performance of the CDSA. The survey finds out that major research studies tried to create a relationship between pivots and non-pivots to gain accuracy. This paper presents a synthesized review of CDSA and explores the current methods and potential future directions. It also addresses the challenges and opportunities presented by these emerging trends and their significance for researchers and practitioners in the CDSA field.

Keywords: ***Cross Domain Sentiment Analysis; Domain Adaptation; PRISMA 2020; Transfer Learning***

1. Introduction

Textual data has become ubiquitous on the internet and provides numerous opportunities for analysis and interpretation. With 5.18 billion people using the internet in 2023[1], we can expect this number to soar to 6 billion by 2028[2]. Humans increasingly rely on social media and the internet as a primary means of connecting with the outside world. People trust online posts, reviews, and tweets for information sharing and accepting the truth behind them.

An ever-increasing amount of data is being generated on the web servers daily, with users generating approximately 402.74 million TB daily[3]. However, this raw user-generated data only becomes valuable when processed to gain actionable information for decision-making. Data Mining and text modeling are gaining success in using digital data in many knowledge engineering areas like classification, clustering, machine translation, entity relationship extraction, prediction [1], and social media analytics. These studies were conducted to extract useful information from blogs, user reviews, posts, and news on social media.

[1] https://www.broadbandsearch.net/blog/internet-statistics
[2] https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/
[3] https://www.explodingtopics.com/blog/data-generated-per-day

Rati Sharma and Kamlesh Lakhwani, "A Systematic Literature Review on Cross Domain Sentiment Analysis Techniques: PRISMA Approach", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 30-55, Vol. 8, No. 4, 1st October 2024, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2024.04.002, Available: http://aetic.theiaer.org/archive/v8/v8n4/p2.html.

This information is heavily utilized by people and institutions for decision-making. The classification task is considered a significant role player in artificial intelligence research, whether by capturing image features [2] or linguistic features

Sentiment analysis (SA) evaluates the emotions or opinions conveyed in the product reviews to classify them as like or dislike or as positive or negative. SA captures the underlying tone behind the reviews [3, 4]. Natural Language Processing (NLP) offers a statistical approach to SA using Machine Learning (ML). The performance of SA approaches relies on higher occurrences of Labeled Data (LDa) within the same domain. For instance, a SA classifier trained on the camera domain will give accurate results if future reviews come only from the camera domain. The classifier built for the Book domain may not provide good results for the baby product domain because the same word and expression may have different polarities in variant domains. It can be seen from below example:

a) Baby product domain - *It is **easy** for her to hold.* - Positive polarity.
b) Book review domain - *The conclusion of this book is **easy** to predict.* - Negative polarity.

The word **easy** conveys different polarity in different domains [5]. This problem is caused by the variation in the data distribution and feature space across cross-domains, called *domain shift* or *domain gap*. When there is a variation in the data distribution, the SA model has to be built from scratch with a new collection of LDa. However, capturing, annotating, computing, and modeling this data repeatedly for different domains involves huge costs. The challenge is to exploit the great volume of the existing LDa across domains, identify the polarity of the Unlabeled Data (ULDa), and treat the situation of *domain shift*. In the last decade, the dominance of literature has aided in addressing the domain shift issue through *domain adaptation*. This whole process is comprehended as Cross Domain Sentiment Analysis (CDSA).

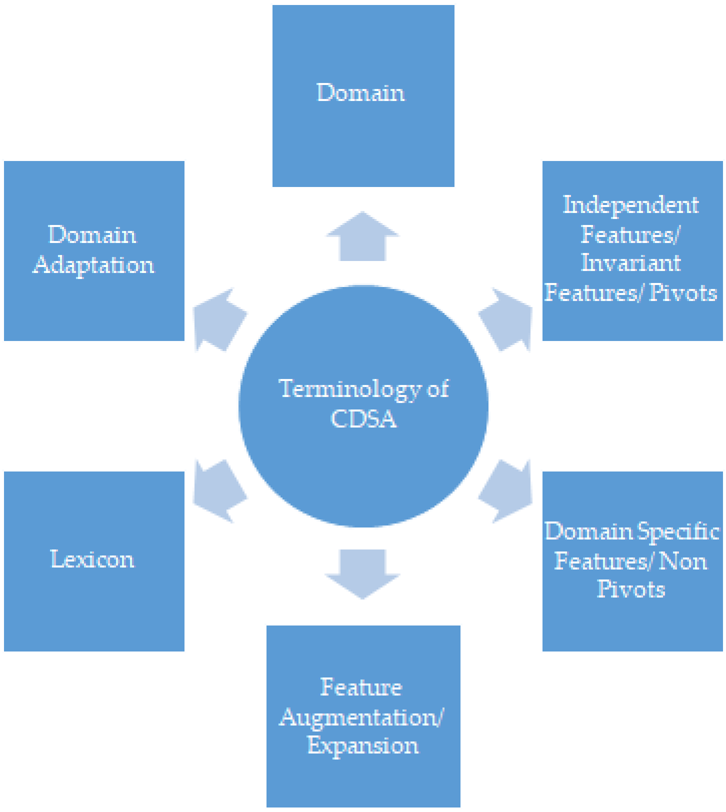

Figure 1. Terminology used in CDSA

2. Important Terminology

The basic terminology used in the research paper is mentioned in Figure 1 and described below:

- **Domain:** Domain is a collection of data having similar characteristics. For example, the camera, baby product, and electronics domain.

- **Domain-Independent Features/ Domain-Shared Features / Common Features/ Invariant Features/ Pivots:** These features occur frequently and carry the same sentiment polarity in all domains. For example, “good”, “fast” and “low”.
- **Domain-Dependent Features/ Domain-Specific Features/ Non-Pivots:** These features convey emotions within a specific field. For example, “soft” is a characteristic of fabric.
- **Domain Adaptation:** This knowledge learning and transferring approach uses LDa present in one domain, namely source domain (SD), using or not some data from another domain, target domain (TD), to annotate unlabeled instances in the target domain automatically.
- **Feature Augmentation/Expansion:** Some features are present in the SD but absent in TD. Feature augmentation is the method of expanding the feature space by adding more related features to it.
- **Lexicon:** Lexicon is a list of words carrying emotions behind them: positive, negative, neutral.

3. Research Questions

The research paper presents a comprehensive survey of research studies published on CDSA from 2017 to 2023. This survey employed baseline methods and aimed to uncover the insights that would answer the following questions:

1) RQ1: How does CDSA impacted by data distribution between SD and TD?
2) RQ2: How does the performance of the CDSA varies with different techniques?
3) RQ3: What are the most relevant parameters considered while performing CDSA?
4) RQ4: Which newly developed methods have been proven to become the baseline methods for comparing CDSA results in the future?

The exploration of the above research questions is crucial for improving the performance and accuracy of CDSA functionality in further research endeavors.

4. Literature Work

CDSA is a prominent topic in the research community in the realm of ML, NLP, image processing, and Data mining. This research paper explores the existing review articles to understand how other authors analyzed the recently developed approaches with the aim of enriching the CDSA, as explained below and depicted in Table 1.

Table 1. Existing review articles of CDSA

Author	Year	No. Of Papers Studied in Detail	Duration	No. Of Approaches Mentioned for CDSA
[6]	2017	28	2010-2016	12
[6]	2017	32	2007-2017	7
[7]	2018	6		3
[8]	2019	28		11
[9]	2020	11	2010-2019	10

A survey [10] is mentioned that studied 28 research articles about CDSA, classified various techniques used in literature into 12 types, and discussed their advantages and disadvantages. Another review was done [6] on 32 research articles in detail from 2007 to 2017, and CDSA approaches were analyzed in terms of techniques used, dataset (DS) used, and accuracy measured. Apart from that, the author found that the accuracy of CDSA lies between 50% and 85%. Three groups (SFA, SCL, JST) [7] are created for CDSA techniques. Techniques used in the reviewed research paper lie in either of the groups. The main issues in CDSA are 1) Sparsity, 2) Polysemy, 3) Feature Divergence, and 4) Polarity Divergence [8]. The author summarized that the performance of the CDSA approaches relies on the LDa in the SD. The author highlights that apart from the problems mentioned above in CDSA approaches, a few more factors need to be taken care of to gain the performance: linguistic diversities, cultural variations, and availability of real-world data sets from different industries. It has been found [9] that SFA is a considerable State-Of-The-Art (SOTA) methodology.

These review articles lack answers to the research questions mentioned in the above section. Non-clarity on the answers from the existing literature reviews on the CDSA led us to perform this literature work, which is different from earlier work regarding systematic review methodology, analysis, findings, and comparison with SOTA methods. The novelties of this literature review are mentioned:

1. The PRISMA 2020 [11] methodology is used for the first time to perform the systematic literature review on CDSA.
2. This is the first study to determine the variation in accuracy of different established techniques for CDSA.

This study is the first to analyze the validation process from the selected studies to get the most favorable conditions so that the best accuracy can be achieved in future research.

5. Review Process

The survey process started with searching databases and digital libraries such as ScienceDirect, SpringerLink, and IEEE Xplore.

5.1. Search Criteria

Taking in mind the research questions, the search process started with keywords related to the CDSA, such as “sentiment”, “cross domain”, “opinion mining”, and their synonyms. Since the CDSA is relatively a new area of research, these keywords provide so many irrelevant studies. So, the search is refined many times to increase the possibility of capturing the most relevant articles. We used queries to search for all relevant studies related to the CDSA from the years 2017 to 2023. The search queries are shown in Figure 2.

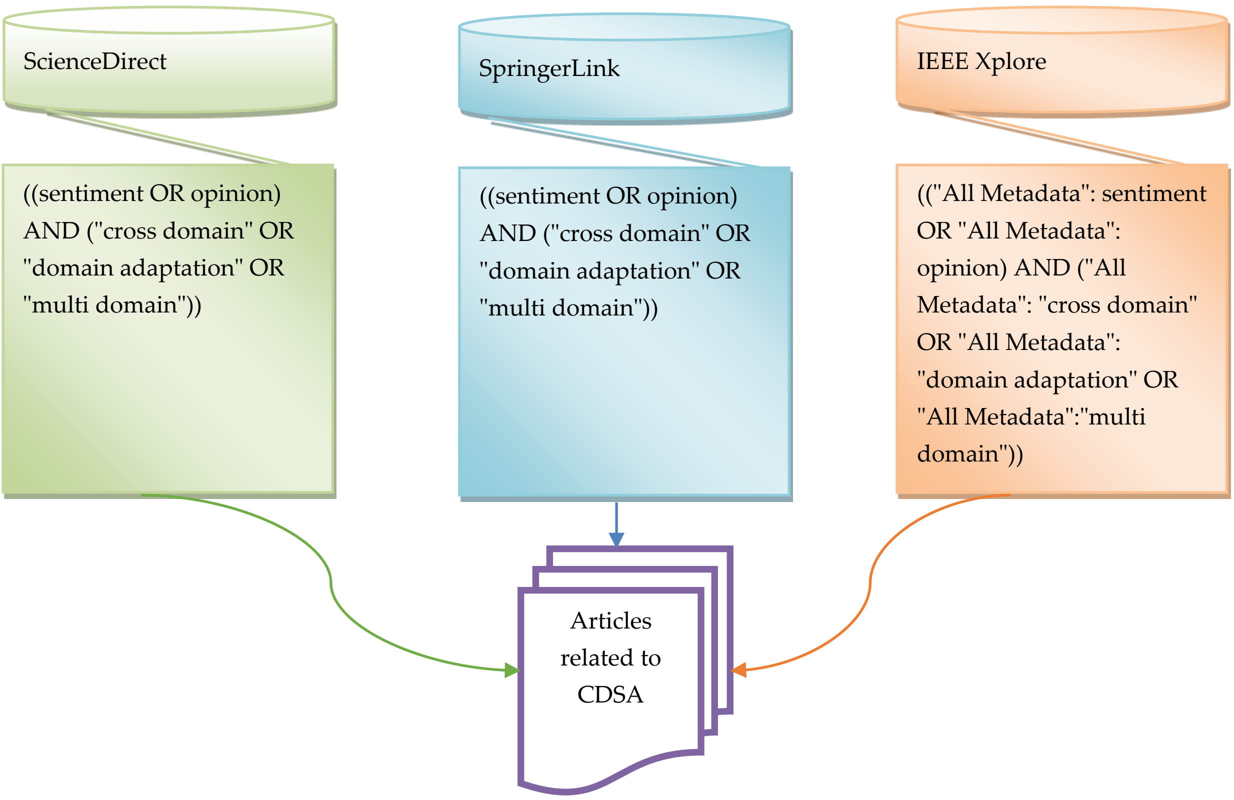

Figure 2. Digital libraries and databases with search queries.

5.2. Review Methodology

The methodology used for systemic review is the PRISMA 2020 model [11] because this model ensures a transparent and full systematic review with meta-analysis. This research is the first time studying the existing literature on CDSA by following the PRISMA 2020 model, as mentioned in Figure 3.

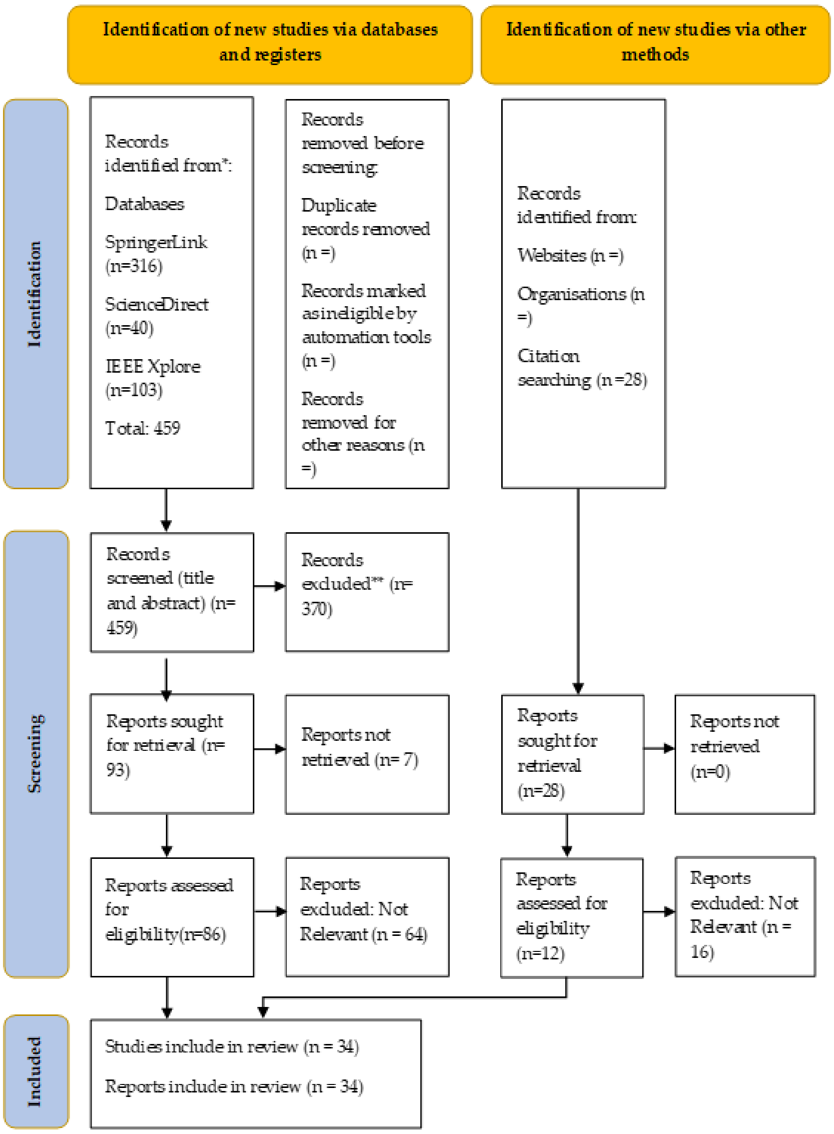

Figure 3. PRISMA 2020 flow diagram for CDSA

5.3. Inclusion and Exclusion Criteria

Some sets of rules are taken into account to include only relevant articles and exclude irrelevant studies from the collection of papers captured from digital libraries and databases. These set of rules, as mentioned in Table 2, will help this study to aligned the papers to fulfill the research objectives.

Table 2. Inclusion and Exclusion Criteria

S. No.	Inclusion Criteria	Exclusion Criteria
1	Date ranges from 2017-2023	Articles that don't focus on CDSA.
2	Reputed journals	Lacks in providing results for proposed method.
3	Reputed conference	Lacks in achieving the mentioned objectives
4	English language articles	Lacks in evaluating performance measures.

The inclusion criteria for this systematic literature review started from the year ranging from 2017 to 2023. Following the inclusion criteria of dates, only those articles that are part of reputed journals and conference proceedings (indexed in SCOPUS, Web of Science, DOAJ, etc.) are included. All the articles are relevant to CDSA only. The exclusion criterion started by going through first the title, abstract, and finally, the full text of each research paper by excluding those not focusing on the CDSA.

6. Extraction and Analysis of Research Studies

A total of 34 papers were chosen for an extensive review, as mentioned in Table 3. The table includes the methodology, validation, and limitations of the recent year's literature.

7. Different techniques of CDSA

The techniques identified after studying the 34 research papers are classified into seven groups, explained below and mentioned in Figure 4.

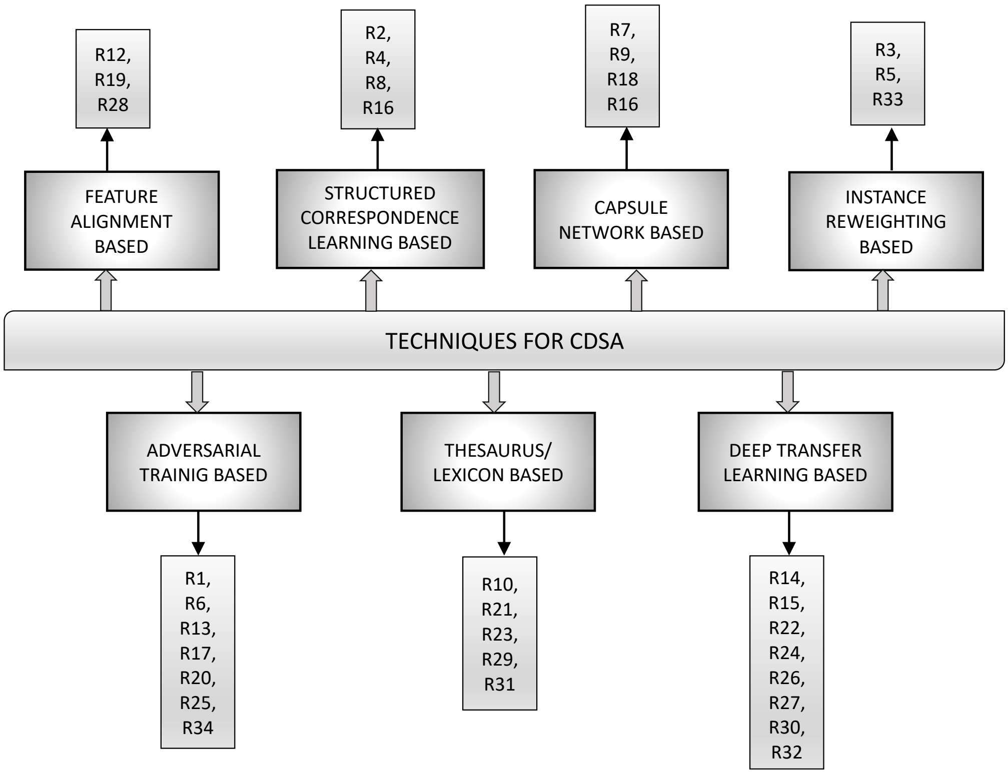

Figure 4. Techniques for CDSA

7.1. Feature Alignment-Based Approaches

The SFA approach was introduced by Pan, Ni, Sun, Yang and Chen [12] to rectify the cross-domain problem in the SA field. The space between both domains was reduced utilizing the Bipartite Graph (BG) and Domain-Independent Features (DIF) while creating clusters of the Domain-Specific Features (DSF) and DIF with the help of a co-occurrence matrix. The SFA algorithm consists of four major components: 1) DIF selection, 2) BG Construction, 3) Spectral Feature Clustering, and 4) Feature Augmentation. The SFA was extended by Jia et al. [13] through learning in-direct mapping rules, which were present in DIF and DSF features with the help of a graph-based Apriori algorithm [14]. The author used the graph to get the indirect mappings of DSF in both domains by taking the help of DIF. These mapping rules have improved the performance of CDSA with a rich feature set to get better accuracy than baseline SFA. Later, a necessary extension of SFA was again attempted [15], for feature extraction along with merging the synonyms and processing the negation words to capture the DIF. TD specific features were captured using BG and proved to have better accuracy than SFA.

On the other hand, the CrossWord algorithm [16] was suggested, where the main idea was to use the stochastic neighbor embedding method [17] and overcome the spectral approaches where embedding space didn't hold desired information accurately because those methods preserved neighbor features poorly. The authors extracted pivots and non-pivots, the most repeatedly co-occurring features in the same domain. Pivots were selected by pivot polarity graph, whereas non-pivots were selected by co-occurrence BG. Stochastic word graph preservation extracts the semantic information of words and creates alignment between the SD and TD. The authors used a stochastic neighbor embedding technique to preserve similarity structures and a mapping function was used to reduce the algorithm bias with single-layer NN.

Table 3. Methodology, validation and limitation of selected research studies

Research Paper	Proposed Method	Technique	Key
		Validation	
		Limitation	
[18]	Adversarial Memory Network (AMN)	The method was created to automatically extract pivots. This model contains four components: word attention, sentiment classifier (SC), domain classifier (DC), and regularizer. The joint learning model shares parameters and visualizes the quality features extracted from both domains.	R1
		The method was validated on 12 cross-domain pairs of Amazon review DS [19] and achieved an average accuracy of 87.21%. Another variant called AMN-NA depends on the LDa in the SD.	
		Lacks in considering non-pivots as they also contain important information.	
[20]	AE-SCL-SR	The author has extended SCL incorporation with Neural Networks (NN), an unsupervised DA, and exploits the power of autoencoder for pivot extraction in CDSA. The author has created two versions of this method: 1) First model autoencoders, and 2) The second model integrates the first method with Similarity Regularization.	R2
		The results were tested on ARD and proved that AE-SCL-SR gives better accuracy for the generalization of CDSA.	
		It avoids structure-aware classification and generates only one representation vector for each input.	
[21]	Prediction Reweighting approach for Domain Adaptation (PRDA)	A novel technique minimized the gap between the SD and TD by a reweighting strategy based on convex optimization techniques [22] and signed distances to increase the adaptability of the classifier.	R3
		It has been validated that the geometry structure and point-wise reweighting scheme make the method more adaptable.	
[23]	Significant Consistent Polarity (SCP)	The new method captured the SCP as transferable knowledge from across domains using the Chi-square test and cosine similarity so that baseline SCL methods can be improved. A cross-domain SC was built, trained on SCP words in the labeled SD, and injected into unlabeled TD to improve the CDSA.	R4
		The results proved that the gold standard SCP words do the CDSA more accurately than the other standard SCL algorithms and provided 85.75 % accuracy.	
		Pivots are selected manually.	
[24]	Domain Specific Feature Transfer (DSFT)	The author has proposed cutoff and fillup strategies combining DSF with domain-in-variant features to enhance the power of homogeneous feature space and make the CDSA incomparable without using auxiliary information. The method utilized Maximum Mean Discrepancy so that the distribution gap in both domains can be reduced and DIF can be mapped with DSF.	R5
		The results validated that DSFT gave a good accuracy of 80.47% and proved that conversion of heterogeneous DA into homogeneous DA increased the accuracy of the CDSA task.	
		Covariate shift exists due to using the original feature span to process the DIF.	
[25]	Hierarchical Attention Transfer Network (HATN	The HATN is a higher variant of AMN and captures the pivots via P-net and non-pivots via NP-net automatically. The HATN training process learns essential information in two modes: individual attention learning and joint attention learning.	R6
		The method is validated on the real-world DS and proved that the pivots and non-pivots are crucial for CDSA. This method pays attention to those sentences that contain non-pivot features and raises the accuracy level.	

		Hierarchical positional encoding has increased the complexity, and the contextual meaning of the sentence is ignored.	
[26]	Transferring Capabilities of Capsule network (TL-Capsule)	The author used a capsule network for the first time to solve the CDSA task. This method compressed the capsule to make the routing faster.	R7
		The ablation study is conducted with some baseline methods, and it is proved that the performance of the capsule network improved by using three strategies: leaky softmax as Activation Function (AF), orphan category, and coefficients amendment strategies for the CDSA task.	
		Lacks in capturing latent features of documents.	
[27]	Pivot Based Language Model (PBLM)	A new model was created to increase the accuracy of CDSA by taking care of 1) NN modelling, 2) Pivot-based modeling, and 3) Structure awareness of the task classifier.	R8
		Two variants of PBLM were experimented on two DS, and the performance of pivot-based models with integration with NN was significantly improved. The results showed that PBLM gave better accuracy when there was a larger difference in domain and data distribution.	
		Selects numerous pivots, which made the model complex and not robust enough to be applied to various domains since hyper parameter and tuning were biased.	
[28]	Capsule network with the Domain Adaptation scenario with semantic Rules (CapsuleDAR)	The author has devised this method for exploiting the part-whole relationships with DIF to fill the gap in both domains. The Base network and Rule network are two significant parts of this method. The base network handles pivot selection, whereas semantic rules are incorporated with the rule networks.	R9
		The experimental results showed the approach preserves the intrinsic correlation in the sentence and fills the gap between the domains with weight initialization and sentence structure.	
		The model is complicated due to iterative routing with poor generalization and visualization.	
[29]	Lexicon-based Method	A lexicon-based method was created that uses SentiWordNet to extract the polarity of features and map by using LDa from both the domains and ULDa from the TD. Stanford POS tagging was used with a maximum entropy classifier to solve the CDSA task.	R10
		The experimental results showed higher accuracy than the baseline methods where book, camera, and movie domains consist of common features and words that are granularized. Results have also shown that using BGs to create clusters has reduced the difference between non-pivots in different domains.	
		External resources add complications.	
[30]	SRI	The author proposed a method that utilized the Sentiment Related Index (SRI) to fill the space between DIF and DSF. The method uses UDa from the TD to power up the unbiasedness of the task.	R11
		The method is validated on two DS and achieved the highest accuracy compared to all other baseline methods, which are biased towards the infrequent feature while considering the shorter text.	
		Multiple SDs are required to achieve a good performance.	
[13]	Words Alignment Based on Association Rules (WAAR)	The algorithm was constructed to reduce the distribution by using cross-domains to create a relationship between DIF with the help of extracted mappings between DIF and DSF.	R12
		The validation proved that using the Apriori algorithm [14] improved the effectiveness of the suggested method as compared to baseline methods.	
		Lacks in dealing with diversified sentiment orientation in pivots.	
[31]	CCHAN	CCHAN was created with the combination of a cloze task network (CTN) and Convolution hierarchical attention networks (CHAN) based on two staged attention mechanisms. CTN was used to extract features, and CHAN classified ULDa of the SD.	R13
		The author created twenty pairs for validation of the method, which improved the test results of existing attention-based CDSA methods. It has been reported that this model can be allied to other domains easily since it doesn't require labeled pivots.	

		Sentiment factors are ignored for extracted features.	
[32]	CNN_FT	The author has created an inductive transfer learning approach using the multi-convolution layers called CNN_FT. The five-layered network structure was the same for both domains, where weights were shared at the convolution layer to transfer the features from the SD to TD using a few LDa from the TD. A drop out layer [33] was added to avoid the overfitting problem, and a maximum pooling layer was introduced so that essential features could be extracted and transferred.	R14
		The method was tested on the famous Amazon DS with supervised ML methods (SVM, NB, and LR) and DANN and proved that CNN_FT gave better results since it did not require manual pivot selection in any case. Another advantage of performance improvement was fine-tuning with the small amount of the labeled TD to reduce the gaps between domains.	
		It is not emphasized that extracting attentive semantic features and context of the text is important.	
[34]	Mutltihead attention method	A new framework was proposed based on a pertained architecture to solve CDSA tasks without a huge amount of LDa using transformers (BERT and XLNet). Multihead attention was used to learn the features that occurred in more than one place. BERT and XLNet are fine-tuned on LDa from the SD and used to identify the class of the ULDa in the TD.	R15
		The results were compared with alternative methods like AMN, HATN, and HANP. It has been found that the BERT and XLNet outperform the SOTA techniques. It was also found that XLNet performed better than BERT in capturing context information.	
		XLNet is more data-hungry.	
[35]	Task Refinement Learning (TRL)	The method was created as an extension to PBLM to make the representation learning more robust by clustering the pivots according to the information they contain. The proposed model predicts the clusters instead of pivots iteratively. The gradual training process of this network gave smoother convergence so that the input sequence structure could be exploited in a more satisfactory way.	R16
		TRL was tested on the Amazon review DS, and validated that the method improved the performance of the representation learning with NN by preserving the information in the cluster of pivots.	
		Two stages are involved in accomplishing the task, which works independently and makes architecture more complex.	
[36]	Hierarchical Attention Network with Prior knowledge information (HANP)	The proposed study has incorporated the dis-pivot selection to the attention network with CNN to improve the performance of CDSA. A hierarchical layer with the attention technique has been incorporated to identify the essential words and sentences so that pivots and non-pivots are extracted at the same time. A Dictionary-based approach (Sentiwordnet 3.0) [37] is used to inject prior knowledge into the NN to increase the accuracy level.	R17
		The model was tested on 20 pairs of Amazon review DS, and prior knowledge positively impacted the CDSA task. The author analysed the accuracy of the CDSA as directly proportional to the count of pivots, non-pivots, and dis-pivots. It has been concluded that HANP has effectively differentiated the polarity of the same feature.	
		Lacks while considering the necessary domain knowledge.	
[38]	Capsule network method with Identifying Transferable Knowledge (CITK)	A new method exploited the part-whole relationship in learning the DSF. The SCP words were used to identify those features, and the model bridged the gap between the SD and TD to make the CDSA more accurate.	R18
		The results showed that CITK improved the model's efficiency even when highly noisy data was present in the DS. The author also concluded that highly compatible domains were the kitchen and the electronics domain.	
		Negative transfer is ignored.	
[16]	CrossWord	The author has boosted the potential of embedding techniques using similarity relationships between the DIF and DSF employing the stochastic neighbor embedding [17] method to preserve the polarity information accurately. The author has focused on the design complexity and explored the simpler mapping method to reduce the computational cost as well as preserve a similar structure accurately.	R19

		Tested on Amazon review database and 12 pairs of cross-domain. Results proved that the stochastic embedding method was superior to spectral embedding methods.	
		Neighborhood contains shorter area of feature.	
[39]	Wasserstein based Transfer Network (WTN).	A new method was constructed based on the Wasserstein Distance (WD) for CDSA. The attention layer captured the essential features to fill the gap between both domains. The author has extracted semantic knowledge from the text using BERT, whereas, feature extraction was done via Gated Recurrent Units (GRU). WD distance was used to reduce the domain discrepancy and extract pivots.	R20
		Results were tested on 12 pairs of Amazon review DS and gave 91 % accuracy. It was also proved that NN based and memory-based systems have improved the performance of CDSA when compared to SOTA methods.	
		Study lacks in exploiting DSF.	
[40]	BCP	The author proposed Bidirectional Conditional Probability (BCP) to measure the unbalanced cooccurred features and create the sentiment-sensitive thesaurus. The author has done a comparative study of PMI, PMIsquare, Gmeans, and EMI.	R21
		Results showed that BCP outperformed other methods, and SVM gave better results than NB. The method was experimented with seven pairs from SD to TD.	
		Negative transfer and noisy lexical elements can be added while feature expansion.	
[41]	Attention Network based on Feature Sequences (ANFS)	A novel method was an improvement of CNN_FT, combined CNN, BiLSTM, and attention mechanism to extract essential information: 1) local semantic features and 2) contextual semantic relationships to provide good quality feature extraction and reduce negative transfer to improve CDSA task.	R22
		The viability of the method says CDSA performance is improved if the model transfers the network (parameter shared) trained by the SD to the TD and fine-tune it with little labeled TD.	
		Results are improved when TD contains 2.5 % LDa.	
[42]	DIL	Domain Independent Lexicon (DIL) was constructed, which used sentiment lexicon for computation of sentence polarity of unlabeled samples in the TD with the help of the multilayer perceptron model. Bing Liu's sentiment lexicon [43], SentiWordNet2 [37] ,and MPQA [44] were used to construct DIL.	R23
		This method proved that taking different vocabularies for a general sentiment lexicon gave better results than taking a single lexicon-based dictionary.	
		Ample training time is required due to the extensive training corpus to achieve accuracy.	
[45]	Deep Transfer Learning Mechanism (DTLM)	The DTLM was introduced to better utilize unlabeled data in the TD and improve the CDSA using BERT and Kullback-Leibler (KL) divergence. Unlabeled samples from the TD were processed via 1) entropy minimization, 2) KL divergence, and 3) consistency regularization to adjust the weights in both domains to capture the domain-invariant features.	R24
		Two DS were used, and performance was tested by analyzing the F-macro and accuracy to verify the viability of the suggested method. It was concluded that KL divergence can effectively map the various features into a unified embedding space to capture the pivots.	
		Lacks in handling diversified expressions.	
[46]	Graph Domain Adversarial Transfer Network (GDATN	An unsupervised CDSA method was developed to classify sentiments and domains via unsupervised learning to lower the data distribution in SD and TD. DC uses the gradient-based backpropagation mechanism to extract the DIF, while SC uses a projection mechanism to identify the DSF.	R25
		Results proved that differences between the domains were reduced with adversarial learning and complex relationships between words were extracted with the graph-based model.	
		A bidirectional mechanism alone was not sufficient to achieve good performance.	
[47]	Parameter Transferring and Attention Sharing Mechanism (PTASM)	The method is related to HATN, where a Hierarchical Attention Network (HAN) was introduced for attention sharing and feature extraction at both levels, word and sentence, to improve the task of CDSA.	R26
		It was validated that word-level parameter sharing gave cutting-edge results compared to sentence-level parameter sharing after testing on 12 pairs of Amazon review DS.	
		Lacks in generalization for CDSA.	

[48]	Sentiment-Sensitive Network Model (SSNM)	A network was constructed to share emotions from pivots and non-pivots. Prior knowledge was utilized from the sentiment dictionary. Two networks were created to fetch pivots and non-pivots with joint attention learning and interpretable information for emotion transfer between both domains. Feature extraction was done with a Bidirectional Gated Recurrent Unit (BiGRU) while empowering the linguistic details and reducing the domain discrepancy with WD.	R27
		Two variants of SSNM (GloVe and BERT) were tested on 20 pairs of Amazon review DS and achieved remarkable compared to comparison to SOTA methods. The author investigated that WD gave better results than GRL.	
		The performance relies on the frequency of sentiment features in the dictionary.	
[15]	Cross-Domain Sentiment Analysis by Refining Feature Extraction (CDSARFE)	A novel method was suggested to identify semantic and syntactic properties to capture similar words from both domains. The feature extraction technique replaced negative polarity words with antonyms to improve the mapping mechanism and enhance the DA. BG reduced the gap between data distributions in both domains.	R28
		The findings proved that the feature-opinion pair resolves the vocabulary mismatch problem and improves the quality of CDSA in comparison to SCL and SFA.	
		Lacks in revealing the relationship between domain invariant and DSF.	
[49]	ELMO	A transfer learning method based on ELMO was developed, which provides accessible plug-and-play parameters to solve CDSA effectively.	R29
		The method was tested on Twitter DS and proved that ELMO can be integrated with NN to achieve better accuracy.	
		Lacks in generalization.	
[50]	Structure Consistent Cross-domain word Embedding (SCE)	A word embedding was devised to use the relationships in pivots by obtaining the domain structure using the semantic graph and co-occurred DSF. The author has used the Laplacian Eigenmaps [51] to maintain consistency in domain structure while performing DA.	R30
		Results proved that this method effectively propagates the semantic information from SD to TD while preserving the domain structure when compared to the SOTA methods.	
		Task-oriented domain structure is not captured.	
[52]	Dual Word Embedding (DWE)	A novel word embedding method was proposed using the BERT semantic channel and word2vec syntactic channel, which captured semantic and syntactic information textCNN and graph attention mechanism. Sentiment classification loss was computed using a cross-entropy function with a SC.	R31
		The method was tested on Amazon review DS, and it proved that using DWE retains the part of the information that was lost using single-word embedding.	
		Applying channels makes the method more complex.	
[53]	Graph Adaptive Semantic Transfer (GAST)	A new method was proposed to extract major linguistic features from the LDa: 1) textual and 2) graph adaptive semantics. The model employs two modules: a) POS Transformer is used to encode the features, and 3) Hybrid Graph Attention (HGAT) is used to weigh the feature and learn the syntactic relations of the sentence. Domain discriminator uses softmax function to classify the sentiments.	R32
		GAST was tested and proved more effective in transferring the features from SD to TD than other baseline methods.	
		Performance degrades when more heads are stacked.	
[54]		The author has introduced a novel CDSA method that relies on reducing the impact of domain shift by increasing the interclass margin with the help of the Gaussian Mixture Model (GMM). The boundaries of different classes are increased to get rid of misclassification. This method has used tf-idf and BERT as feature extractors.	R33
		The results were promising when margins between classes were increased compared to SOTA methods.	
		An initial large gap in between different domains can misalign some classes.	

[54]	Feature Projection and Multi-source Attention (FPMA)	FPMA lowered the impact of negative transfer in CDSA by employing two feature extractors and one domain discriminator. The power of feature representation is improved by capturing the DSF using an adversarial network and then optimizing with orthogonal projection.	R34
		The method was tested and proved to be improved over the baseline method by targeting multiple TDs at one time, which benefits IOT researchers by providing more effectiveness.	
		Data-hungry algorithm.	

7.2. Structured Correspondence Learning (SCL) Based Approaches

The Structured Correspondence Learning algorithm was created [56] to handle the CDSA situation with supervised learning. This method works in 4 steps: 1) Pivot features are identified from the ULDa from the SD and TD. 2) Pivot predictors are created with each pivot to identify correlated non-pivot with a high degree of correspondence. 3) Feature vector augmentation is done by applying a mapping. 3) a standard discriminative learner is used to train the SD so that it can perform well on the TD. The work was extended [19] by selecting pivots with mutual information to train the classifier. SCL has been again extended by Ziser and Richard [20] with incorporation with NN to exploit the power of autoencoder for pivot extraction in CDSA. This method encodes the non-pivots into a hidden layer, and the decoder will give the correlated pivots based on the decoding matrix. Two variants were created: 1) With autoencoder SCL (AE-SCL) and 2) Along with similarity regularization (AE-SCL-SR). AE-SCL-SR used pre-embedded pivots, where similar-meaning pivots have the same vectors. Another variant of this model used pre-trained embeddings for pivot features. The model was trained with word2vec [57] embeddings and representation learning using ULDa from the SD and TD.

The current learning methods avoid structure-aware classification. This limitation was overcome in the following year [27] with PBLM, an unsupervised DA method, where the structure of input text was taken care of with a structure-aware classifier. Two variants of this method were created, PBLM-LSTM and PBLM-CNN, and tested on two DS and outperformed on challenging DS (product to airline and airline to product), where data distribution and domain discrepancy were larger than other setups. The results proved that the PBLM improved the efficiency of pivot-based learning when the model was integrated with NN. The performance of LSTM-based approaches relies on the hyperparameter tuning as well as the design of the NN [58]. On the other hand, the extraction of features from ULDa from the SD and TD gives numerous pivots, which tend to increase the complexity of the classifier and decrease the quality of CDSA. To solve this problem, the author proposed Task Refinement Learning (TRL) [35], which iteratively trained the PBLM model and progressively revealed more details about every pivot. The clusters of pivots were made, and the classifier predicted those clusters iteratively by ranking them according to the information they conveyed. The ranking of pivots was done using three methods: MI, Frequency and, Similar Frequencies.

Another attempt to improve the SCL was made by Sharma, Bhattacharyya, Dandapat [23], and Bhatt [23] for the extraction of the Significant Consistent Polarity (SCP) features from both domains utilizing two statistical methods. The first statistical method is the Chi-square test, used to extract SCP features, known as gold standard words. The gold standard words are those words that were important in both the domains, and polarity remains unchanged. The second statistical method is cosine similarity, with the help of which the polarity of unlabeled occurrences in the TD is identified. Unlike the BG constructed in SCL, the author used cosine similarity to identify pivots more effectively. Results proved that the SCP words used by this approach were more accurately determined than the other SCL methods.

7.3. Thesaurus / Lexicon Based Approaches

This is another important technique for creating and using thesaurus to solve the CDSA. Bollegala, Weir and Carroll [59] created Sentiment Sensitive Thesaurus (SST) for the features that exhibit the same meaning in different domains and achieved better accuracy than SCL and SFA. SST was extended as ESST [60] with the help of wiktionary.

But this early research was poorly performed when applied to short text and overcome by Wang, Niu, Song and Atiquzzaman [30] with a method that was dedicated to short text based on the Sentiment Related

Index (SRI) as shown in equation 1 and 2, which was calculated to lower the distribution gap between the features from both domains. The authors used HowNet6 and NTUSD to get the general DIF set to extract the DSF and classify the reviews in the TD.

$$\mathrm{SRI}\ (\mathrm{s},\mathrm{t}) = \frac{1}{\Sigma \mathrm{w} \in \mathrm{V} \mathrm{dist}(\mathrm{w},\mathrm{s},\mathrm{t})} \tag{1}$$

$$\mathrm{dist}\ (\mathrm{w,s,t}) = \begin{cases} P(w|R_t).\log(\frac{P(w|R_t)}{P(w|(R_s \ \cup \ R_t))}) & if\ w \in V_{s,t} \\ 0 & otherwise, \end{cases} \tag{2}$$

where, R is the product review set, R_s and R_t are reviews where DIF and candidate features are present in R. V is the vocabulary set and $V_{s,t}$ is the features set that are present in $R_s \cup R_t$. The author used word occurrence distribution to capture the SRI instead of word co-occurrence, like in PMI-based methods, to improve performance when short text reviews exist.

A novel approach, BCP, was proposed [40], with the main objective of pointing out the most informative and representative transferable features by generating thesaurus automatically. The co-occurrence calculations of features were done with five methods, as shown in equations 3, 4, 5, 6, and 7, respectively.

$$\text{Pointwise Mutual Information PMI (x, y)} = \log \frac{\mathrm{P(x,y)}}{\mathrm{P(x)P(y)}} \tag{3}$$

$$\text{Pointwise Mutual Information Square } (\mathrm{PMI}^2) = \log \frac{\mathrm{P(x,y)}^2}{\mathrm{P(x)P(y)}} \tag{4}$$

$$\text{Gmean (x, y)} = \frac{\mathrm{f(x,y)}}{\sqrt{\mathrm{f(x)f(y)}}} \tag{5}$$

$$\text{Enhanced Mutual Information EMI (x, y)} = \log \frac{\mathrm{P(x,y)}}{(\mathrm{P(x)} - \mathrm{P(x,y)})(\mathrm{P(y)} - \mathrm{P(x,y)})} \tag{6}$$

$$\text{Bidirectional Conditional Probability BCP (x, y)} = \frac{\mathrm{P(x,y)}\{\mathrm{P(y)} + \mathrm{P(x)}\}}{(\mathrm{P(x)} - \mathrm{P(x,y)})(\mathrm{P(y} - \mathrm{P(x,y)}))} \tag{7}$$

where P(x) and P(y) are the probabilities of occurring features x and respectively, and P (x, y) is the probability of co-occurring features x and y. The author introduced BCP to calculate the unsymmetrical co-occurrence features, which outperformed the other four methods.

7.4. Instance Reweighting Based Approaches

These are another line of approaches to solving the DA problem where, the SD classifier could not be directly used for predict of the polarity of the TD because of the difference in the data distribution. Li, Song and Huan [21] proposed a method called the Prediction Reweighting approach for Domain Adaptation (PRDA), where a domain separator was used to separate the data distribution of both domains. PRDA calculates the signed distance of the TD instances to check its closeness from the domain separator and reweights each instance based on its closeness. This closeness was also used as a confidence measure of correct predictions of the TD instances with the SD classifier. Later on, Wei, Ke and Goh [24] tried to resolve the domain discrepancy problem underlying the CDSA and proposed Domain Specific Feature Transfer (DSFT). This method is implemented in two ways: 1) The "Cut-off", where the feature space is aligned with the help of DIF only, and 2) The "Fill-Up" method utilizes DSF to enhance the common feature space to extract the discriminative power of the DSF. The authors brought the two different domains closer by reweighting shared features with the help of DSF sets. Along with checking the closeness of the two domains with NDCG, the PRDA was validated, and accuracy was proved on the famous Amazon DS.

7.5. Adversarial Training-Based Approaches

An approach Domain Adversarial Neural Network (DANN) [61] was suggested as a semi-supervised approach with improvement on marginalized Stacked Autoencoders (mSDA) [62] where a few data was labeled in the TD. DANN is a feed-forward NN that can be applied to the TD by learning the mappings between both domains, the gradient reversal layer used for backpropagation. DANN was built mainly of three components: 1) Label Predictor, 2) DC, and 3) Feature Extractor. Li, Zhang, Wei, Wu and Yang [18] incorporated a memory network into DANN and proposed AMN, where an attention mechanism was used to improve the automatic interpretability and extraction of pivots. The main aim of the attention mechanism was to identify essential words for the SA. The AMN is a jointly trained model that contains four components: 1) Word attention mechanism has been developed to capture the essential features automatically, 2) SC classifies the documents in an unsupervised manner and minimizes the cross entropy while predicting the sentiment label, 3) DC increases the model's capacity to discriminate the SD and TD, and 4) Regularization handles the overfitting in both classifiers. The whole model was jointly optimized by backpropagation and focused on the word level attention.

AMN didn't take care of DSF and was considered in the following year by Wei, Zhang and Yang [25]. The author developed a Hierarchical Attention Transfer Network (HATN) and extended the AMN to automatically predict the DIF and DSF. The proposed method exploits the hierarchical structure of the document. The significant parts of HATN involve two networks: 1) P-net and 2) NP-net. The former network identifies pivots, which are domain-shared features, and the latter network identifies non-pivots, which are DSF. HATN projects DSF into DIF space. The training process of HATN involves two phases: 1) Individual Attention Learning enforces both networks to work separately, and 2) Joint Attention Learning concatenates the data representation achieved by both networks for CDSA. The complexity of HATN was higher due to the labeling of the pivots, which increases annotation errors. Manshu and Xuemin [31] solved these complexities of HATN in a newly developed method, a combination of CTN and CHAN, abbreviated as CCHAN. CTN extracted essential words according to occurrence and sentences in both domains using two layered BiLSTM. CCHAN is a three-layered CNN network used to classify class labels. The authors used an attention mechanism in two stages to identify important words and sentences. Both networks jointly learn the information from the SD and TD to train and validate the model on TD. The result claimed that CCHAN worked effectively when the author used a dictionary. The study has proved that hierarchical structure was better than non-hierarchical models and extraction of n-gram feature lights positively on CDSA.

Later, HANP [36][36] was introduced to improve the power of CDSA by incorporating dis-pivot selection with pivots and non-pivots in the domains. The authors used three components to build this model:1) Sentiment Dictionary Match (SDM), 2) CNN, and 3) The Hierarchical attention network. The authors have injected prior knowledge into the NN. The sentimental information was captured by SDM with the help of SentiWordNet3.0 [37], and contextual information was captured by CNN. SDM creates a connection with the help of 1) features from TD - non-pivots, dis-pivots, and 2) features from the SD-pivots, non-pivots. VDCNN [63] is famous for transfer learning, so the author used it in HANP. Du, He, Wang and Zhang [39] constructed a new method based on WD for CDSA named Wasserstein-based Transfer Network (WTN). WTN works in four modules: 1) Encoding unit based on BERT and word2vec. 2) Feature extraction is done with a BiGRU to obtain important domain invariant features. 3) Domain discrepancy learning unit checks the WD between distributions of the SD and TD. The lesser the distance, the more DIF can be extracted. 4) The classification module classifies the TD data with joint learning capability. Results proved that WTN improved the performance of CDSA significantly and is more stable.

Recently, Tang, Mi, Xue and Cao [46] developed GDATN, which works in an unsupervised manner. GDATN has two parts: SC and DC. Both networks represent the features using BiLSTM and attention network. The main idea of the authors behind GDATN was to find out the hidden and shared relationship between the SD and TD to obtain better features from both domains. The gradient reversal layer was applied to extract DIF features for domain classification. A projection mechanism was used to project the selected features into the target feature space for better prediction.

7.6. Deep Transfer Learning-Based Approaches

DTL is an approach utilizing deep learning models where knowledge transfer is done from the labeled SD to a few or unlabeled TD. It has been taken into account that NN has dominated the medical image processing domain [60], but NN has also received similar importance in NLP. [64] Meng, Long, Yu, Zhao and Liu [65] proposed a model transfer-based approach CNN_FT using CNN for better feature extraction with BOW and a skip-gram method to capture rich semantic knowledge for better CDSA. CNN_FT is a five-layered method where LDa from the SD is utilized for model training purposes. After training, the weights and parameters are shared with the TD model, which contains a structure similar to that of the SD. Then, the TD model uses some labeled samples from TD, fine-tuned in the last layer using the stochastic descendent method. CNN_FT suffers from incomplete feature extraction, which was solved by Meng, Dong, Long and Zhao [41] while developing ANFS. The ANFS combines CNN, BiLSTM, and an attention mechanism, which reduces negative transfer and provides good-quality feature extraction to improve CDSA. In ANFS, CNN extracted local feature, BiLSTM preserved long-term dependency, and the attention mechanism captures the important words. The author's study mentioned that ANFS is a six-layered architecture, where two layers (BiLSTM and attention) are fine-tuned to get the updated information extracted from the attention mechanism to improve the results for classification.

Recently, PTASM [47] was created, which is related to HATN in capturing word and sentence-level information. The attention mechanism was incorporated into NN in PTASM. PTASM was based on transferable NN, where SD and TD have similar network structures, and parameters are shared between the networks. In ANFS, both the networks are five layered Hierarchical Attention Network (HAN) with Gate Recurrent Units (GRU), discovering important features for sentiment classification. The authors described in their study that ANFS is a model-parameter transfer network where a parameter-transferring mechanism was introduced at word or sentence level between two domains. The performance of PTASM mainly depends on two factors: 1) Parameter identification and 2) Parameter transfer strategy. PTASM gave higher accuracy at world-level transfer than sentence-level transfer. After that, SSNM [48] was constructed to transfer emotions across domains in the form of attention. The study mentioned that the features were extracted with two novel methods: 1) KPE-net was used to fetch the DIF, and 2) NKPE-net was used to fetch the DSF. A Bidirectional Gated Recurrent Unit (BiGRU) has been utilized to fetch features before classifying sentiments. While developing SSNM, the authors constructed a WD-based domain confusion component using pivots to shorten the distribution gap between both domains. Two variants of SSNM (GloVe and BERT) were tested on Amazon review DS and achieved remarkable performance compared to SOTA methods.

Later, a novel approach, the Deep Transfer Learning Mechanism (DTLM) [45], was proposed to exploit the usage of ULDa from the TD and raise the performance of CDSA. The authors emphasized three main factors of DTLM: BERT, KL divergence, entropy minimization, and consistency regularization. In DTLM, BERT was used as a feature encoder to get emotional polarity between sentences. KL divergence was utilized to make the model domain adaptive using matching score between both domains. Apart from that, DTLM used entropy minimization and consistency regularization to increase the usage of ULDa in TD and improve prediction results.

7.7. Capsule Network-Based Approaches

The capsule network is a group of neurons. The max pool layer in CNN captured the values of the most prominent neurons and shared those values to the following network. In this process, the model losses spatial knowledge of those left behind features. In a capsule network, the spatial information about the low-level features is encoded by the active capsule. It passes to the next-level neurons to learn the part-whole relationships [66]. The CapsuleDAR [28] was created to take care of the part-whole relationship between both domains. The CapsuleDAR consists of two networks that are similar in structure. The first network called the base network, consists of a capsule network where the embedding layer and convolution layer take care of pivot extraction. The second network, named the rule network, is responsible for adding semantic information to the overall architecture to learn the part-whole relationship.

CapsuleDAR is complicated and doesn't take good care of the CDSA's generalization ability. In the following year, an investigation [26] is done on the Transferring Capabilities of Capsule network (TL-Capsule) for CDSA. The semantic representation of embedding was constructed from the high-level features captured with CNN. TL-Capsule does the DA task by adapting the SD to the TD iteratively. Since the capsule is a group of neurons, it takes up lots of memory space. Therefore, a capsule compression layer is added to efficiently use the memory. To improve the capsule networks in the field of CDSA, SCP was incorporated for feature extraction with capsule networks [38] in CITK to learn and transfer part-whole relationships better. SCP features are extracted using BERT. Correlation between the part and the whole is captured by the capsule network.

8. Analysis of Different Parameters for CDSA

This section provides a thorough investigation of the significant factors listed in Table 4, which are considered important while performing the CDSA.

Table 4. Detailed analysis of various parameter for CDSA

Reference	Method	Features		Activation Functions			Labeled		Joint Learning
		Pivot	Non-Pivot	Softmax	Sigmoid	Relu	SD	TD	
[18]	AMN	✓	×	✓			✓	×	✓
[20]	AE-SCL-SR	✓	✓		✓		✓	×	
[21]	PRDA	✓	×				✓	×	
[23]		✓	✓				✓	×	
[24]	DSFT	✓	✓				✓	×	
[25]	HATN	✓	✓	✓			✓	×	✓
[26]	TL-Capsule	✓	×	✓		✓	✓	×	
[27]	PBLM	✓	✓				✓	×	✓
[28]	CapsuleDAR	✓	×	✓			✓	×	
[29]		✓	✓				✓	✓	
[30]	SRI	✓	✓				✓	×	
[13]	WAAR	✓	✓				✓	×	
[31]	CCHAN	✓	✓	✓		✓	✓	×	✓
[32]	CNN_FT			✓			✓	✓	
[34]		✓	×	✓		✓	✓	×	
[35]	TRL	✓	✓	✓			✓	×	
[36]	HANP	✓	✓	✓		✓	✓	×	✓
[38]	CITK	✓	×	✓			✓	×	✓
[16]	CrossWord	✓	✓		✓		✓	×	
[39]	WTN	✓	×	✓			✓	×	✓
[40]		✓					✓	×	
[41]	ANFS			✓		✓	✓	✓	
[42]	DIL						✓	×	
[45]	DTLM			✓			✓	✓	
[46]	GDATN	✓	✓	✓		✓	✓	×	
[47]	PTASM	✓	✓	✓			✓	✓	✓
[48]	SSNM	✓	✓	✓			✓	×	✓
[15]	CDSARFE	✓	✓				✓	✓	
[49]				✓			✓	×	
[50]	SCE	✓	✓				✓	×	✓

8.1. Pivots / Non-Pivots

It has been observed that all of the selected research studies, as shown in Table 4, indicate that the main focus is on the selection of pivots from both domains. As mentioned in Figure 5, 55 % of research studies used pivot features, 36% used both pivots and non-pivot features, and 9 % didn't use any.

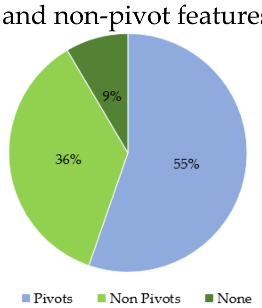

Figure 5. Frequency of selection of pivots, non-pivots in the research studies

8.2. Activation Function / Transfer Function

The AF utilizes the values of the NN and assigns them in the range of 0 to 1. As mentioned in Figure 6., Softmax is the heavily used AF following Sigmoid and Relu in the selected studies. We can see in Table 4 that some research studies used a combination of Softmax and Relu to achieve better results.

8.3. Labeled Data

The aforementioned study in Table 4 indicates that some researchers increased the accuracy of CDSA elegantly in a semi-supervised manner through a few labeled occurrences present in the TD at the time of training. Major research studies chose to solve the challenge behind CDSA where LDa in TD is absent in an unsupervised manner by considering only LDa from the SD. Though semi-supervised approaches give better accuracy, they require manual intervention.

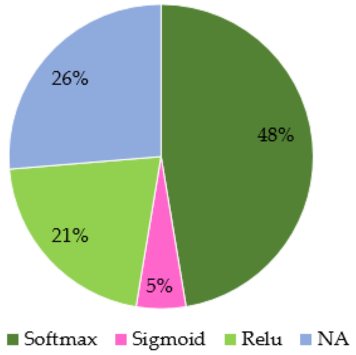

Figure 6. Frequency of the AF used in the selected studies.

8.4. Word Embeddings

NLP uses words as vectors to form word embeddings for statistical computations while doing text analysis. The selected studies used the below-mentioned word embeddings to perform the CDSA task, and Table 5 describes the inclusion of these word embeddings in the respective methods.

- **Word2Vec:** It is the most popular method developed at Google [67] to create numerical representations of word features in the form of vectors. Word2vec provides two model architectures using cosine similarity:
    - **Continuous Bag-of-Words (CBOW):** The method uses projection weights to identify the feature in consideration based on the neighboring context words without considering the order of the context words.
    - **Skip-Gram (SG):** implies that the neighboring context features can be predicted using the intended features. Word2vec is a scalable embedding.
- **Global Vectors (GloVe):** This method was proposed [68] at Stanford University after the great success of Word2Vec. This is based on the matrix factorization method.
- **Bidirectional Encoder Representations from Transformers (BERT):** BERT [69] utilizes previous tokens to make a prediction of the next word [70]. This method works with two phases:1) Masked Language Modeling to identify the next word and 2) Next Sentence Prediction model to predict the next sentence.
- **XLNet:** This language model was developed by Google team members and Carnegie Mellon University [71], which is Auto-Regressive. This model is used for generative tasks, which makes two-way predictions about the next word depending upon the context of the word.

Table 5. Word Embeddings used by selected studies

Word Embedding	Developed At	Used In
Word2vec	Google, US	R1, R2, R4, R6, R7, R8, R9, R13, R14, R16, R17, R20, R22, R25, R30, R31
GloVe	Stanford University	R19, R26, R27, R29, R30, R31, R32
BERT	Google US	R15, R18, R20, R24, R26, R27, R31, R33, R32, R33, R34
XLNet	Carnegie Mellon University	R15

8.5. Lexicons

Lexicons are well-structured collections of lexicons where words are associated with their semantic polarity orientation such that they are positive or negative with scores [72]. There are various sentiment lexicons used in the selected studies, as mentioned in Table 6, and are described below:

- **WordNet:** It is a lexical structured database where words express their distinct concept [73]. Words are present in the WordNet with lexical and semantic connections in the form of synset.
- **SentiWordNet:** Since WordNet is a generalized lexicon and is not useful in the case of SA, SentiWordNet lexicon is devised for SA and contains a polarity score for each WordNet synset [74].
- **SentiWordNet 3.0:** The higher variant of SentiWordNet, created in the year 2010, where sentiment lexicons [37] are created with an improved WordNet database.
- **Multi-Perspective Question Answering (MPQA):** It is a glossary of opinions in the form of a lexicon created at the University of Pittsburg [44] which contains the contextual polarity for subjective expressions (emotion, opinion, evaluation, stance, etc.).

Table 6: Sentiment lexicons used by selected studies

Lexicon Name	Created At	Research Studies
WordNet	Princeton University	R28
SentiWordNet	ISTI, Italy	R10
SentiWordNet 3.0	ISTI, Italy	R17
MPQA	University of Pittsburg	R23

8.6. Baseline Methods Used for Comparison

This section presents the methods that have become baseline methods used by researchers to compare the results obtained in the research done on CDSA. Table 7 describes the most used baseline methods in the selected studies. It has been found that early methods SCL and SFA, along with new methods DANN, DMSDA, AMN, and HATN, are the most used baseline approaches in the selected research papers.

Table 7. SOTA approaches used in the studies

Method Name	Aut hor	Year	Research Studies
SCL	[19]	2006	R1, R7, R11, R12, R R19, R20, R22, R25, R28, R32
SCL-MI	[19]	2007	R2, R7, R8, R14, R18, R22
SFA	[12]	2010	R1, R6, R10, R11, R12, R19, R20, R22, R28, R32
mSDA	[62]	2012	R8, R9, R16, R18, R19, R32
DACNN	[75]	2014	R1, R7, R9, R18, R34
CNN-aux	[76]	2016	R6, R13, R15, R17, R27
DANN	[61]	2016	R1, R6, R7, R9, R14, R18, R19, R20, R22, R25, R27, R32, R34
DMSDA	[61]	2016	R1, R2, R6, R7, R8, R15, R16, R27
AMN	[18]	2017	R6, R9, R13, R15, R17, R20, R25, R27, R32
AE-SCL-SR	[20]	2017	R8, R16, R18, R22
HATN	[25]	2018	R9, R13, R15, R17, R19, R20, R25, R26, R27, R33
CapsuleDAR	[28]	2018	R9, R18, R26, R32
PBLM	[27]	2018	R7, R9, R18, R33
IATN	[77]	2019	R9, R26, R27, R32

8.7. Accuracy

It has been observed that most methods gave accuracy between 80 % and 90 % while solving the problem of CDSA. As shown in Figure 7, only a few of the research studies have reached above 90 %. The highest accuracy achieved was 95.13% [34] by using XLNet.

8.8. MM and NN Models

This critical parameter is the backbone of CDSA. Figure 8 shows the proportion of ML and NN used in the selected research studies. It is clearly visible that NN-based approaches dominate (72%) the CDSA research. LibLinear (LR) is the least used method. 8 % of research studies (R1, R6, R25) had applied adversarial training to use the poorly LDa with less human intervention. The attention mechanism [78] uses a translation model to capture valuable features (R15, R20, R26). Some research studies have paid attention to long-term dependency on LSTM (R8, R16, R17, R22, R25). CNN is heavily used for feature extraction (R7, R8, R9, R13, R14, R16, R17, R18, R22, R27, R29).

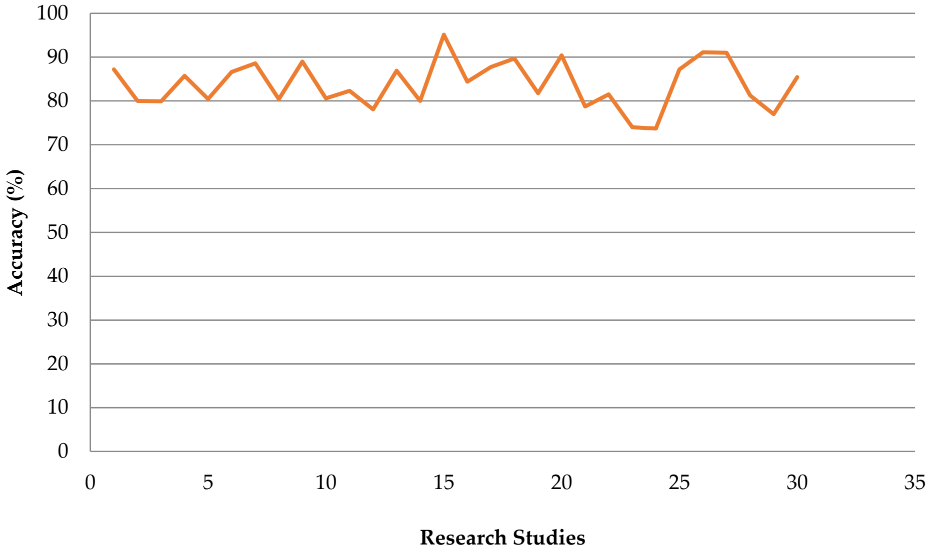

Figure 7. Accuracy achieved in selected studies

The capsule network maintains the semantic compositionality constituting the part-whole relationship between the local and global features (R7, R9, R18).

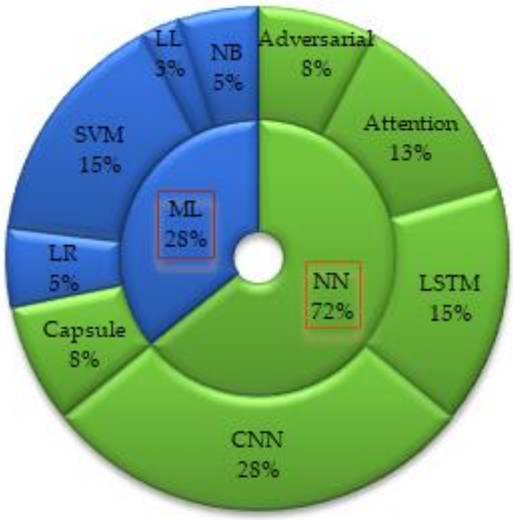

Figure 8. Frequency of ML and NN approaches used in the selected research studies

9. Conclusion

This study offers an analysis of CDSA techniques based on the research studies conducted in the recent past. Firstly, the paper categorized 34 research studies into seven classes and explained them accordingly. It is observed that all classes employ either ML or NN as artificial intelligence methods for CDSA. Secondly, a summary has been presented showing innovative methodologies, validation, and constraints encountered in the selected research studies. Past researchers used Amazon DS heavily to test their proposed methods. The best accuracy score was 95.13 % XLNet, but this method proved to be data-hungry. Thirdly, an analysis was conducted to probe various parameters that impact the performance of CDSA. Word2Vec was found to be the significant word embedding method, closely followed by GloVe and BERT. Softmax has played a legitimate role in CDSA as an SF used in NN approaches. Despite standing as an early-stage research area, SCL substantiated its position as one of the baseline methods, along with AMN and DANN, for comparison of accuracies of new research. Past studies have shown that the main attraction was to find out pivots and non-pivots. SFA was introduced with BG, forming clusters of diverse domain features with the help of a co-occurrence matrix. The Apriori algorithm, Crossword algorithm, and stochastic neighbor algorithm improved the complexities of SFA. With the introduction of SCL, the ability of pivot identifier, pivot predictor, feature vector augmentation, and discriminative learner with supervised learning has improved. Subsequent efforts were directed towards extending SCP to SCP-MI and incorporating the NN with SCP using statistical methods (chi-square test, cosine similarity) to overcome the feature complexity challenges of SCP. Furthermore, variants of PBLM, language modeling with LSTM and CNN, have effectively addressed structure aware classification. PBLM was iteratively refined by TRL to gain more information about the pivotal features. However, the distribution gaps were always there in SFA and SCL. Lexicon-based approaches were introduced to lower the distribution gap between the SD and TD. In this study, we realized that all classes have used either ML or NN as an artificial intelligence technique to perform CDSA.

This paper surveys various methods that have been proposed in the past acknowledging the absence of a perfect solution for addressing CDSA. Despite these advancements, this paper presents our vision to improve the lacking area of CDSA by keeping in mind to

- Use the attention mechanism with adversarial training.
- Find out some measures to check the relatedness of the SD and the TD.
- Transferable and related knowledge should only be shared to reduce the negative transfer.
- Refine the interdisciplinary collaboration.
- Improve the generalizability and scalability considering the diverse domain.

References

[1] Ziauddin Ursani and Ahsan Ahmad Ursani, “The Theory of Probabilistic Hierarchical Learning for Classification”, *Annals of Emerging Technologies in Computing (AETiC),* Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 61-74, Vol. 7, No. 1, 1st January 2023, Published by International Association of Educators and Researchers (IAER), DOI: 10.33166/AETiC.2023.01.005, Available: http://aetic.theiaer.org/archive/v7/v7n1/p5.html.

[2] Nizamuddin Khan, Ajay Singh and Rajeev Agrawal, “Enhancing feature extraction technique through Spatial Deep Learning model for Facial Emotion Detection”, *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, Vol. 7, No. 2, pp. 9-22, 1st April 2023, Published by International Association of Educators and Researchers (IAER), DOI: 10.33166/AETiC.2023.02.002, Available: http://aetic.theiaer.org/archive/v7/v7n2/p2.html.

[3] Kamlesh Lakhwani and Jeevan Bala, “Deep learning for sentiment analysis”, *Journal of Emerging Technologies and Innovative Research (JETIR),* Online ISSN: 2349-5162, Vol. 5, No. 10, pp. 578–584, October 2018, Published by IJ Publication, Available: https://www.jetir.org/papers/JETIRDQ06084.

[4] Kamal Kant Hiran, Ritesh Kumar Jain, Kamlesh Lakhwani and Ruchi Doshi, *Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples*, English ed., Noida, India: BPB Publications, 16th September 2021, ISBN: 9789391392352, Available: https://books.google.co.in/books?id=2_lAzwEACAAJ.

[5] Yitao Cai and Xiaojun Wan, “Multi-domain sentiment classification based on domain-aware embedding and attention”, in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)*, 10-16 August 2019, Macao, China, E- ISBN: 9780999241141, Vol. 2019, pp. 4904–4910, Published by International Joint Conferences on Artificial Intelligence Organisation, DOI: 10.24963/ijcai.2019/681, Available: https://www.ijcai.org/proceedings/2019/681.

[6] Tareq Al-Moslmi, Nazlia Omar, Salwani Abdullah and Mohammed Albared, “Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review”, *IEEE Access*, Online ISSN: 2169-3536, Vol. 5, pp. 16173–16192, 31st March 2017, Published by IEEE, DOI: 10.1109/ACCESS.2017.2690342, Available: https://ieeexplore.ieee.org/document/7891035.

[7] Ashwini Save and Narendra Shekokar, “Analysis of cross domain sentiment techniques”, in *Proceedings of the International Conference on Electrical, Electronics, Communication Computer Technologies and Optimization Techniques, (ICEECCOT 2017)*, 15-16 December 2017, Mysore, India, Print ISBN: 978-1-5386-1205-7, E-ISBN: 978-1-5386-2361-9, pp. 1-9, Published by IEEE, DOI: 10.1109/ICEECCOT.2017.8284637, Available: https://ieeexplore.ieee.org/document/8284637.

[8] Manimekalai V and Gomathi S Rohini, "A Survey on Cross - Domain Opinion Mining", *International Journal of Computer Sciences and Engineering*, Online ISSN: 2347-2693, Vol. 6, No. 10, pp. 792–796, 31st October 2018, Published by International Journal of Computer Sciences and Engineering, DOI: 10.26438/ijcse/v6i10.792796, Available: https://www.ijcseonline.org/full_paper_view.php?paper_id=3102.

[9] Parvati Kadli and Vidyavathi B.M., "Cross Domain Sentiment Classification Techniques: A Review", *International Journal of Computer Applications (IJCA)*, Online ISSN: 0975-8887, Vol. 181, No. 37, pp. 13–20, January 2019, Published by Foundation of Computer Science (FCS), NY, USA, DOI: 10.5120/ijca2019918338, Available: https://www.ijcaonline.org/archives/volume181/number37/30273-2019918338/.

[10] Nancy Kansal, Lipika Goel and Sonam Gupta, "A Literature Review on Cross Domain Sentiment Analysis Using Machine learning", *International Journal of Artificial Intelligence and Machine Learning*, Print ISSN: 2642-1577, Online ISSN: 2642-1585, Vol. 10, No. 2, pp. 43–56, 12th June 2021, Published by IGI Global, DOI: 10.4018/ijaiml.2020070103, Available: https://www.igi-global.com/gateway/article/257271.

[11] Matthew James Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews", *The BMJ*, Online ISSN: 1756-1833, Vol. 372, No. 71, 29th March 2021, Published by BMJ Publishing Group Ltd, DOI: 10.1136/bmj.n71, Available: https://www.bmj.com/content/372/bmj.n71.

[12] Sinno Jialin Pan, Xiaochuan Ni, Jian Tao Sun, Qiang Yang and Zheng Chen, "Cross-domain sentiment classification via spectral feature alignment", in *Proceedings of the 19th International Conference on World Wide Web (WWW 2010)*, 26-30 April 2010, Raleigh, North Caroline, USA, ISBN: 9781605587998, pp. 751–760, Published by Association for Computing Machinery, DOI: 10.1145/1772690.1772767, Available: https://dl.acm.org/doi/10.1145/1772690.1772767.

[13] Xi Bin Jia, Ya Jin, Ning Li, Xing Su, Barry Cardiff *et al.*, "Words alignment based on association rules for cross-domain sentiment classification", *Frontiers of Information Technology and Electronic Engineering*, Print ISSN: 2095-9230, Online ISSN: 2095-9184, Vol. 19, No. 2, pp. 260–272, 19th April 2018, DOI: 10.1631/FITEE.1601679, Available: https://link.springer.com/article/10.1631/FITEE.1601679.

[14] Rakesh Agrawal and Ramakrishnan Srikant, "Fast Algorithms for Mining Association Rules in Large Databases" in *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB 1994)*, 12-15 September 1994, Santiago, Chile, ISBN: 1558601538, pp. 487–499, Published by Morgan Kaufmann Publishers Inc., DOI: 10.5555/645920.672836, Available: https://www.vldb.org/conf/1994/P487.PDF.

[15] Geethapriya A and Valli S, "An Enhanced Approach to Map Domain-Specific Words in Cross-Domain Sentiment Analysis", *Information Systems Frontiers*, Print ISSN: 1387-3326, Online ISSN: 1572-9419, Vol. 23, No. 3, pp. 791–805, 5th January 2021, Published by Kluwer Academic Publishers, United States, DOI: 10.1007/s10796-020-10094-5, Available: https://link.springer.com/article/10.1007/s10796-020-10094-5.

[16] Yanbin Hao, Tingting Mu, Richang Hong, Meng Wang, Xueliang Liu *et al.* , "Cross-Domain Sentiment Encoding through Stochastic Word Embedding", *IEEE Transactions on Knowledge and Data Engineering*, Print ISSN: 1041-4347, Online ISSN: 1558-2191, Vol. 32, No. 10, pp. 1909–1922, 1st October 2020, Published by IEEE, DOI: 10.1109/TKDE.2019.2913379, Available: https://ieeexplore.ieee.org/document/8700271.

[17] Geoffrey Hinton and Sam Roweis, "Stochastic neighbor embedding", in *Proceedings of the 15th International Conference on Neural Information Processing Systems (NIPS 2003)*, 8-13 December 2003, Vancouver and Whistler, British Columbia, Canada, ISBN: 9780262025508, Vol. 15, pp. 857–864, Published by MIT Press, Available: https://dl.acm.org/doi/10.5555/2968618.2968725.

[18] Zheng Li, Yu Zhang, Ying Wei, Yuxiang Wu and Qiang Yang, "End-to-end adversarial memory network for cross-domain sentiment classification", in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2017)*, 19-25 August 2017, Melbourne, Australia, E-ISBN: 978-0-9992411-0-3, pp. 2237–2243, Published by AAAI Press, DOI: 10.24963/ijcai.2017/311, Available: https://www.ijcai.org/proceedings/2017/0311.pdf.

[19] John Blitzer, Ryan McDonald and Fernando Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification", in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 23-30 June 2007, Prague, Czech Republic, pp. 440–447, Published by Association for Computational Linguistics, Available: https://aclanthology.org/P07-1056.pdf.

[20] Yftah Ziser and Roi Reichart, "Neural structural correspondence learning for domain adaptation", in *Proceedings of the 21st Annual conference on Computational Natural Language Learning (CoNLL 2017)*, 3-4 August 2017, Vancouver, Canada, ISBN: 9781945626548, Vol. 1, pp. 400–410, Published by Association for Computational Linguistics, DOI: 10.18653/v1/k17-1040, Available: https://aclanthology.org/K17-1040.

[21] Shuang Li, Shiji Song and Gao Huang, "Prediction reweighting for domain adaptation", *IEEE Transactions on Neural Networks and Learning Systems*, Print ISSN: 2162-237X, Online ISSN: 2162-2388, Vol. 28, No. 7, pp. 1682–1695, April 2017, Published by IEEE, DOI: 10.1109/TNNLS.2016.2538282, Available: https://ieeexplore.ieee.org/document/7457281.

[22] Dimitris Bertsimas and Ioana Popescu, "Optimal inequalities in probability theory: A convex optimization approach", *SIAM Journal on Optimization*, Print ISSN: 1052-6234, Online ISSN: 1095-7189, Vol. 15, No. 3, pp. 780–804, 8th April 2005, Published by Society for Industrial and Applied Mathematics Publications, US, DOI: 10.1137/S1052623401399903, Available: https://epubs.siam.org/doi/10.1137/S1052623401399903.

[23] Raksha Sharma, Pushpak Bhattacharyya, Sandipan Dandapat and Himanshu Sharad Bhatt, "Identifying transferable information across domains for cross-domain sentiment classification", in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 15-20 July 2018, Melbourne Convention and Exhibition Centre, Melbourne, Australia, ISBN: 9781510867093, E-ISBN: 9781510867079, Vol. 1, pp. 968–978, Published by Association for Computational Linguistics, DOI: 10.18653/v1/p18-1089, Available: https://aclanthology.org/P18-1089.

[24] Pengfei Wei, Yiping Ke and Chi K. Goh, "A General Domain Specific Feature Transfer Framework for Hybrid Domain Adaptation", *IEEE Transactions on Knowledge and Data Engineering*, Print ISSN: 1041-4347, Online ISSN: 1558-2191, Vol. 31, No. 8, pp. 1440–1451, 1st August 2019, Published by IEEE, DOI: 10.1109/TKDE.2018.2864732, Available: https://ieeexplore.ieee.org/document/8432087.

[25] Zheng Li, Ying Wei, Yu Zhang and Qiang Yang, "Hierarchical Attention Transfer Network for Cross-Domain Sentiment Classification", in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)*, 2-7 February 2018, Hilton New Orleans Riverside, New Orleans, Louisiana, USA, Print ISSN: 2159-5399, Online ISSN: 2374-3468, Print ISBN: 978-1-57735-800-8, pp. 5852–5859, Published by AAAI Press, Palo Alto, California USA, DOI: 10.1007/978-3-319-93803-5_36, Available: https://cdn.aaai.org/ojs/12055/12055-13-15583-1-2-20201228.pdf.

[26] Min Yang, Wei Zhao, Lei Chen, Qiang Qu, Zhou Zhao *et al.*, "Investigating the transferring capability of capsule networks for text classification", *Neural Networks*, Print ISSN: 0893-6080, Online ISSN: 1879-2782, Vol. 118, pp. 247–261, October 2019, Published by Elsevier Science Ltd. United Kingdom, DOI: 10.1016/j.neunet.2019.06.014, Available: https://www.sciencedirect.com/science/article/abs/pii/S089360801930187X.

[27] Yftah Ziser and Roi Reichart, "Pivot based language modeling for improved neural domain adaptation", in *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Long Papers)*, 1-6 June 2018, New Orleans, Louisiana, USA, Print ISBN: 9781510863507, Vol. 1, pp. 1241–1251, Published by Association for Computational Linguistics, DOI: 10.18653/v1/n18-1112, Available: https://aclanthology.org/N18-1112.pdf.

[28] Bowen Zhang, Xiaofei Xu, Min Yang, Xiaojun Chen and Yunming Ye, "Cross-domain sentiment classification by capsule network with semantic rules", *IEEE Access*, Online ISSN: 2169-3536, Vol. 6, pp. 58284–58294, 10th October 2018, Published by IEEE, DOI: 10.1109/ACCESS.2018.2874623, Available: https://ieeexplore.ieee.org/document/8488682.

[29] Jyoti S. Deshmukh and Amiya Kumar Tripathy, "Entropy based classifier for cross-domain opinion mining", *Applied Computing and Informatics*, Print ISSN: 2634-1964, Online ISSN: 2210-8327, Vol. 14, No. 1, pp. 55–64, January 2018, Published by Elsevier B.V., DOI: 10.1016/j.aci.2017.03.001, Available: https://www.sciencedirect.com/science/article/pii/S2210832717300960.

[30] Lei Wang, Jianwei Niu, Houbing Song and Mohammed Atiquzzaman, "SentiRelated: A cross-domain sentiment classification algorithm for short texts through sentiment related index", *Journal of Network and Computer Applications*, Print ISSN: 1084-8045, Online ISSN: 1095-8592, Vol. 101, pp. 111–119, 1st January 2018, Published by Academic Press Ltd, UK, DOI: 10.1016/j.jnca.2017.11.001, Available: https://www.sciencedirect.com/science/article/abs/pii/S1084804517303582.

[31] Tu Manshu and Zhao Xuemin, "CCHAN: An End-to-End Model for Cross Domain Sentiment Classification", *IEEE Access*, Online ISSN: 2169-3536, Vol. 7, pp. 50232–50239, 14th April 2019, Published by IEEE, DOI: 10.1109/ACCESS.2019.2910300, Available: https://ieeexplore.ieee.org/document/8691665.

[32] Jiana Meng, Yingchun Long, Yuhai Yu, Dandan Zhao and Shuang Liu, "Cross-domain text sentiment analysis based on CNN_FT method", *Information*, Online ISSN: 2078-2489, Vol. 10, No. 5, 25th February 2019, Published by MDPI, DOI: 10.3390/info10050162, Available: https://www.mdpi.com/2078-2489/10/5/162.

[33] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *Journal of Machine Learning Research*, Print ISSN: 1532-4435, Online ISSN: 1532-7928, Vol. 15, No. 56, pp. 1929–1958, July 2014, Published by Microtome Publishing, Available: https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf.

[34] Batsergelen Myagmar, Jie Li and Shigetomo Kimura, "Cross-Domain Sentiment Classification with Bidirectional Contextualized Transformer Language Models", *IEEE Access*, Online ISSN: 2169-3536, Vol. 7, pp. 163219–163230, 8th November 2019, Published by IEEE, DOI: 10.1109/ACCESS.2019.2952360, Available: https://ieeexplore.ieee.org/document/8894409.

[35] Yftah Ziser and Roi Reichart, "Task refinement learning for improved accuracy and stability of unsupervised domain adaptation", in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019),* 28 July – 2 August 2019, Florence, Italy, ISBN: 9781510863507, pp. 5895–5906, Published by Association for Computational Linguistics, DOI: 10.18653/v1/p19-1591, Available: https://aclanthology.org/P19-1591.pdf.

[36] Tu Manshu and Wang Bing, "Adding Prior Knowledge in Hierarchical Attention Neural Network for Cross Domain Sentiment Classification", *IEEE Access*, Online ISSN: 2169-3536, Vol. 7, pp. 32578–32588, 4th March 2019, Published by IEEE, DOI: 10.1109/ACCESS.2019.2901929, Available: https://ieeexplore.ieee.org/document/8658231.

[37] Stefano Baccianella Andrea Esuli and Fabrizio Sebastiani, "SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining", in *Proceedings of the 7th International Conference on Language Resources and Evaluation, (LREC 2010)*, 17-23 May 2010, Valletta, Malta, Print ISBN: 2-9517408-6-7, pp. 2200–2204, Published by European Language Resources Association (ELRA), Available: http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf.

[38] Hongxia Yin, Peiyu Liu, Zhenfang Zhu, Wenkuan Li and Qianqian Wang, "Capsule Network with Identifying Transferable Knowledge for Cross-Domain Sentiment Classification", *IEEE Access*, Online ISSN: 2169-3536, Vol. 7, pp. 153171–153182, 21st October 2019, Published by IEEE, DOI: 10.1109/ACCESS.2019.2948628, Available: https://ieeexplore.ieee.org/document/8658231.

[39] Yongping Du, Meng He, Lulin Wang and Haitong Zhang, "Wasserstein based transfer network for cross-domain sentiment classification", *Knowledge-Based Systems*, Print ISSN: 1872-7409, Online ISSN: 0950-7051, Vol. 204, pp. 106162, 27th September 2020, Published by Elsevier B.V, DOI: 10.1016/j.knosys.2020.106162, Available: https://www.sciencedirect.com/science/article/pii/S0950705120304032.

[40] Tareq Al-Moslmi, Mohammed Albared, Adel Al-Shabi, Salwani Abdullah and Nazlia Omar, "A Comparative Study of Co-Occurrence Strategies for Building A Cross-Domain Sentiment Thesaurus", in *Proceedings of the 1st International Conference of Intelligent Computing and Engineering (ICOICE 2019)*, 15-16 December 2019, Hadhramout, Yemen, Print ISBN:978-1-7281-4488-7, Online ISBN: 978-1-7281-4487-0, pp. 1–8, Published by IEEE, DOI: 10.1109/ICOICE48418.2019.9035179, Available: https://ieeexplore.ieee.org/document/9035179.

[41] Jiana Meng, Yu Dong, Yingchun Long and Dandan Zhao, "An attention network based on feature sequences for cross-domain sentiment classification", *Intelligent Data Analysis*, Print ISSN: 1088-467X, Online ISSN: 1571-4128, Vol. 25, No. 3, pp. 627–640, 20th April 2021, Published by IOS Press, DOI: 10.3233/IDA-205130, Available: https://content.iospress.com/articles/intelligent-data-analysis/ida205130.

[42] Omid Mohamad Beigi and Mohammad Hossein Moattar, "Automatic construction of domain-specific sentiment lexicon for unsupervised domain adaptation and sentiment classification", *Knowledge-Based Systems*, Print ISSN: 0950-7051, Online ISSN: 1872-7409, Vol. 213, p. 106423, 15th February 2021, DOI: 10.1016/j.knosys.2020.106423, Available: https://www.sciencedirect.com/science/article/abs/pii/S0950705120305529.

[43] Minqing Hu and Bing Liu, "Mining and summarizing customer reviews", in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004)*, 22-25 August 2004, Seattle, Washington, USA, Print ISBN: 9781581138887, pp. 168–177, Published by Association for Computing Machinery, New York, USA, DOI: 10.1145/1014052.1014073, Available: https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf.

[44] Theresa Wilson, Janyce Wiebe and Paul Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis", in *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT 2005),* 6-8 October 2005, Vancouver, British Columbia, Canada, pp. 347–354, Published by Association for Computational Linguistics, USA, DOI: 10.3115/1220575.1220619, Available: https://aclanthology.org/H05-1044.pdf.

[45] Zixuan Cao, Yongmei Zhou, Aimin Yang and Sancheng Peng, "Deep transfer learning mechanism for fine-grained cross-domain sentiment classification", *Connection Science*, Print ISSN: 0954-0091, Online ISSN: 1360-0494, Vol. 33, No. 4, pp. 911–928, 17th May 2021, Published by Taylor and Francis, DOI: 10.1080/09540091.2021.1912711, Available: https://www.tandfonline.com/doi/full/10.1080/09540091.2021.1912711.

[46] H. Tang, Y. Mi, F. Xue and Y. Cao, "Graph Domain Adversarial Transfer Network for Cross-Domain Sentiment Classification", *IEEE Access*, Online ISSN: 2169-3536, Vol. 9, pp. 33051–33060, 22nd February 2021, Published by IEEE, DOI: 10.1109/ACCESS.2021.3061139, Available: https://ieeexplore.ieee.org/document/9360543.

[47] Chuanjun Zhao, Suge Wang, Deyu Li, Xianzhi Liu, Xinyi Yang *et al.*, "Cross-domain sentiment classification via parameter transferring and attention sharing mechanism", *Information Sciences*, Print ISSN: 0020-0255, Online ISSN: 1872-6291, Vol. 578, pp. 281–296, November 2021, Published by Elsevier Sciences Inc., DOI: 10.1016/j.ins.2021.07.001, Available: https://www.sciencedirect.com/science/article/abs/pii/S0020025521006915.

[48] Yanping Fu and Yun Liu, "Cross-domain sentiment classification based on key pivot and non-pivot extraction", *Knowledge-Based Systems*, Print ISSN: 0950-7051, Online ISSN: 1872-7409, Vol. 228, p. 107280, 27th September 2021,

Published by Elsevier B.V, DOI: 10.1016/j.knosys.2021.107280, Available: https://www.sciencedirect.com/science/article/abs/pii/S0950705121005426.

[49] Vamshi B. Krishna, Ajeet Kumar Pandey and Siva Kumar A. P., "Universally domain adaptive algorithm for sentiment classification using transfer learning approach", *International Journal of System Assurance Engineering and Management*, Print ISSN: 0975-6809, Online ISSN: 0976-4348, Vol. 12, No. 3, pp. 542–552, 8th May 2021, Published by Springer India, DOI: 10.1007/s13198-021-01113-y, Available: https://link.springer.com/article/10.1007/s13198-021-01113-y.

[50] Heyan Huang and Qian Liu, "Domain structure-based transfer learning for cross-domain word representation", *Information Fusion*, Print ISSN: 1566-2535, Online ISSN: 1872-6305, Vol. 76, pp. 145–156, December 2021, Published by Elsevier B.V, DOI: 10.1016/j.inffus.2021.05.013, Available: https://www.sciencedirect.com/science/article/abs/pii/S1566253521001135.

[51] Mikhail Belkin and Partha Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation", *Neural Computation*, Print ISSN: 0899-7667, Online ISSN: 1530-888X, Vol. 15, No. 6, pp. 1373–1396, 1st June 2003, Published by MIT Press, DOI: 10.1162/089976603321780317, Available: https://www2.imm.dtu.dk/projects/manifold/Papers/Laplacian.pdf.

[52] Zihao Lu, Xiaohui Hu and Yun Xue, "Dual-Word Embedding Model Considering Syntactic Information for Cross-Domain Sentiment Classification", *Mathematics*, Online ISSN: 2227-7390, Vol. 10, No. 24, p. 4704, 11th December 2022, Published by MDPI, DOI: 10.3390/math10244704, Available: https://www.mdpi.com/2227-7390/10/24/4704.

[53] Kai Zhang, Qi Liu, Zhenya Huang, Mingyue Cheng, Kun Zhang *et al.*, "Graph Adaptive Semantic Transfer for Cross-domain Sentiment Classification", in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2022)*, 11-15 July 2022, Madrid, Spain, ISBN: 9781450387323, pp. 1566–1576, Published by Association for Computing Machinery, DOI: 10.1145/3477495.3531984, Available: https://dl.acm.org/doi/pdf/10.1145/3477495.3531984.

[54] Mohammad Rostami, Digbalay Bose, Shrikanth Narayanan and Aram Galstyan, "Domain Adaptation for Sentiment Analysis Using Robust Internal Representations", in *Proceedings of the Findings of the Association for Computational Linguistics(EMNLP 2023)*, 6-10 December 2023, Singapore, ISBN: 9781713885931, pp. 11484–11498, Published by Association for Computational Linguistics, DOI: 10.18653/v1/2023.findings-emnlp.769, Available: https://aclanthology.org/2023.findings-emnlp.769.pdf.

[55] Yeqiu Kong, Zhongwei Xu and Meng Mei, "Cross-Domain Sentiment Analysis Based on Feature Projection and Multi-Source Attention in IoT", *Sensors*, Online ISSN: 1424-8220, Vol. 23, No. 16, p. 7282, 20th August 2023, Published by MDPI, DOI: 10.3390/s23167282, Available: https://www.mdpi.com/1424-8220/23/16/7282.

[56] John Blitzer, Ryan McDonald and Fernando Pereira, "Domain adaptation with structural correspondence learning", in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, 22-23 July 2006, Sydney, Australia, Print ISBN: 1932432736, pp. 120–128, Published by Association for Computational Linguistics, USA, DOI: 10.3115/1610075.1610094, Available: https://aclanthology.org/W06-1615.pdf.

[57] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean, "Distributed representations of words and phrases and their compositionality", in *Proceedings of the Conference on Neural Information Processing Systems (NIPS 2013)*, 5-10 December 2013, Lake Tahoe, Nevada, USA, Print ISBN: 9781632660244, Vol. 26, pp. 1–9, Published by Curran Associates Inc., DOI: 10.48550/arXiv.1310.4546, Available: https://papers.nips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf .

[58] Nils Reimers and Iryna Gurevych, "Reporting Score Distributions Makes a Difference: Performance Study of LSTM-networks for Sequence Tagging ", in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, 7-11 September, Copenhagen, Denmark, Print ISBN: 978-1-945626-83-8, pp. 338–348, Published by Association for Computational Linguistics, DOI: 10.18653/v1/D17-1035, Available: https://aclanthology.org/D17-1035.pdf .

[59] Danushka Bollegala, David. Weir and John Carroll, "Cross-domain sentiment classification using a sentiment sensitive thesaurus", *IEEE Transactions on Knowledge Data Engineering*, Print ISSN: 1041-4347, Online ISSN: 1558-2191, Vol. 25, No. 8, pp. 1719–1731, August 2013, Published by IEEE, DOI: 10.1109/TKDE.2012.103, Available: https://ieeexplore.ieee.org/document/6203505 .

[60] P Sanju and T T. Mirnalinee, "Construction of Enhanced Sentiment Sensitive Thesaurus for Cross Domain Sentiment Classification Using Wiktionary", in *Proceedings of the 3rd International Conference on Soft Computing for Problem Solving (SocProcs 2013). Advances in Intelligent Systems and Computing*, 26-28 December 2013, IIT Roorkee, India, Print ISBN: 978-81-322-1767-1, Online ISBN: 978-81-322-1768-8, Vol. 259, pp. 195-206, Published by Springer India, DOI: 10.1007/978-81-322-1768-8, Available: https://link.springer.com/chapter/10.1007/978-81-322-1768-8_18.

[61] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle *et al.*, "Domain-adversarial training of neural networks", in Advances in Computer Vision and Pattern Recognition: Domain Adaptation in

Computer Vision Applications, Singapore: Springer Nature, 2016, Print ISBN: 978-3-319-58346-4, Online ISBN: 978-3-319-58347-1, Vol. 17, Ch. 10, pp. 189–209, Published by Springer Verlag, DOI: 10.1007/978-3-319-58347-1_10, Available: https://link.springer.com/chapter/10.1007/978-3-319-58347-1_10.

[62] Minmin Chen, Zhixiang Eddie Xu, Kilian Q Weinberger, St Louis and Fei sha, "Marginalized Denoising Autoencoders for Domain Adaptation", in *Proceedings of the 29th International Conference on Machine Learning (ICML 2012),* 26 June-1 July 2012, Edinburgh, Scotland, UK, Print ISBN: 9781450312851, pp. 627–1634, Published by Omnipress Madison, DOI: 10.5555/3042573.3042781, Available: https://icml.cc/2012/papers/416.pdf .

[63] Jason Yosinski, Jeff Clune, Yoshua Bengio and Hod Lipson, "How transferable are features in deep neural networks?", in *Proceedings of the 27th International Conference on Neural Information Processing Systems, (NIPS 2014)*, 8-13 December 2014, Montreal, Canada, Print ISBN: 9781510800410, Vol. 27, pp. 3320–3328, Published by Curran Associates Inc., DOI: 10.48550/arXiv.1411.1792, Available: https://dl.acm.org/doi/10.5555/2969033.2969197.

[64] J.Andrew Onesimu, Varun Unnikrishnan Nair, Martin K. Sagayam, Jennifer Eunice, Mohd Helmy abd Wahab *et al.*, "SkCanNet: A Deep Learning based Skin Cancer Classification Approach", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516- 0281, Online ISSN: 2516-029X, Vol. 7, No. 4, pp. 35-45, 1st October 2023, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2023.04.004, Available: http://aetic.theiaer.org/archive/v7/v7n4/p4.html.

[65] Jiana Meng, Yingchun Long, Yuhai Yu, Dandan Zhao and Shuang Liu, "Cross-domain text sentiment analysis based on CNN_FT method", *Information*, Online ISSN**:** 2078-2489, Vol. 10, No. 5, p. 162, 1st May 2019, Published by MDPI, DOI: 10.3390/info10050162, Available: https://www.mdpi.com/2078-2489/10/5/162 .

[66] Sara Sabour, Nicholas Frosst and Geoffrey E. Hinton, "Dynamic Routing Between Capsules", in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*, 4-9 December 2017, Long Beach, CA, USA, Print ISBN: 9781510860964, Vol. 22, No. 4, pp. 185–186, Published by Curran Associates Inc., DOI: 10.48550/arXiv.1710.09829, Available: https://dl.acm.org/doi/10.5555/3294996.3295142 .

[67] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean, "Efficient estimation of word representations in vector space", in *Proceedings of the International Conference on Learning Representations 2013 (ICLR '13)*, 2-4 May 2013, Scottsdale, Arizona, USA, pp. 1301–3781, DOI: 10.48550/arXiv.1301.3781, Available: https://arxiv.org/abs/1301.3781.

[68] Jeffrey Pennington, Richard Socher and Christopher D. Manning, "GloVe: Global vectors for word representation", in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, 25-29 October 2014, Doha, Qatar, Print ISBN: 978-1-937284-96-1, pp. 1532–1543, Published by Association for Computational Linguistics, DOI: 10.3115/v1/d14-1162, Available: http://www.aclweb.org/anthology/D14-1162 .

[69] Jacob Devlin, Ming Wei Chang, Kenton Lee and Kristina Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding", in *Proceedings of the 2019 Conference of North American Chapter of the Association for Computational Linguistics*, 2-7 June 2019, Minnesota, Minneapolis, USA, Print ISBN: 978-1-950737-13-0, Vol. 1, pp. 4171–4186, Published by Association for Computational Linguistics, DOI: 10.18653/v1/N19-1423, Available: https://aclanthology.org/N19-1423/.

[70] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston and Rob Fergus, "End-to-end memory networks", in *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2015)*, 7-12 December 2015, Montreal, Canada, Print ISBN: 9781510825024, Vol. 28, pp. 2440–2448, Published by Curran Associates Inc., DOI: 10.48550/arXiv.1503.08895, Available: https://dl.acm.org/doi/10.5555/2969442.2969512.

[71] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov *et al.*, "XLNet: Generalized autoregressive pretraining for language understanding", in *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NIPS 2019)*, 8-14 December 2019, Vancouver, Canada, Print ISBN: 9781713807933, Vol. 32, pp. 5733–5763, Published by Curran Associates Inc., DOI: 10.48550/arXiv.1906.08237, Available: https://dl.acm.org/doi/10.5555/3454287.3454804.

[72] Anna Jurek, Maurice D. Mulvenna and Yaxin Bi, "Improved lexicon-based sentiment analysis for social media analytics", *Security Informatics*, Online ISSN: 2190-8532, Vol. 4, No. 9, pp. 1-13, 9th December 2015, Published by SpringerOpen, DOI: 10.1186/s13388-015-0024-x, Available: https://security-informatics.springeropen.com/articles/10.1186/s13388-015-0024-x .

[73] George Armitage Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross and Katherine J. Miller, "Introduction to wordnet: An on-line lexical database", *International Journal of Lexicography*, Print ISSN: 0950-3846, Online ISSN: 1477-4577, Vol. 3, No. 4, pp. 235–244, 1st December 1990, Published by Oxford University Press, DOI: 10.1093/ijl/3.4.235, Available: https://academic.oup.com/ijl/article-abstract/3/4/235/923280 .

[74] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani, "SENTIWORDNET: A publicly available lexical resource for opinion mining", in *Proceedings of the 5th International Conference on Language Resources and Evaluation, (LREC 2006)*, 24-26 May 2006, Genoa, Italy, Print ISBN: 2-9517408-2-4, Online ISBN: 978-2-9517408-2-2, pp. 417–422,

Published by European Language Resources Association (ELRA), Available: http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf.

[75] Yoon Kim, "Convolutional neural networks for sentence classification", in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, (EMNLP 2014)*, 25-29 October 2014, Doha, Qatar, ISBN: 9781634394789, pp. 1746–1751, Published by Association for Computational Linguistics (ACL), DOI: 10.3115/v1/d14-1181, Available: https://aclanthology.org/D14-1181.pdf.

[76] Jianfei Yu and Jing Jiang, "Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification", in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)*, 1-4 November 2016, Austin, Texas, USA, ISBN: 978-1-945626-25-8, pp. 236–246, Published by Association for Computational Linguistics (ACL), DOI: 10.18653/v1/d16-1023, Available: https://aclanthology.org/D16-1023.pdf.

[77] Kai Zhang, Hefu Zhang, Qi Liu, Hongke Zhao, Hengshu Zhu *et al.*, "Interactive attention transfer network for cross-domain sentiment classification", in *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI 2019)*, 27 January-1 February 2019, Honolulu, Hawaii, USA, ISBN: 978-1-57735-809-1, pp. 5773–5780, Published by Association for the Advancement of Artificial Intelligence, DOI: 10.1609/aaai.v33i01.33015773, Available: https://cdn.aaai.org/ojs/4524/4524-13-7563-1-10-20190706.pdf.

[78] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio, "Neural machine translation by jointly learning to align and translate", in *Proceedings of the 3rd International Conference on Learning Representations, (ICLR 2015)*, 7-9 May 2015, San, Diego, CA, USA, pp. 1–15, Published by OpenReview.net, DOI: 10.48550/arXiv.1409.0473, Available: https://arxiv.org/pdf/1409.0473.