Journal of Advanced Research Design 138, Issue 1 (2026) 129-136

Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Evaluating Cross-Domain Sentiment Analysis using Convolutional Neural Network for Amazon Dataset

Azwa Abdul Aziz[1,2,*], Afiq Nasri Othman[1], Pascal Ezenkwu[3], Ellisa Nadia Madi[1,2]

[1] Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Tembila Campus, 22200 Besut, Terengganu, Malaysia
[2] Data Science & Analytics (DASA), Special Interest Group, Universiti Sultan Zainal Abidin (UniSZA), 21300 Kuala Nerus, Terengganu, Malaysia
[3] School of Computing, Engineering & Technology, Robert Gordon University, AB10 7AQ Aberdeen, United Kingdom

ARTICLE INFO

Article history:
Received 3 February 2025
Received in revised form 17 March 2025
Accepted 30 June 2025
Available online 20 July 2025

Keywords:
Sentiment analysis; deep learning; convolutional neural network; cross-domain analysis

ABSTRACT

Sentiment analysis (SA) has garnered extensive research attention over the past decades as a means to comprehend users' attitudes and opinions in various domains. With the proliferation of online communities and the rapid generation of social media content, understanding sentiments has become crucial for decision-makers and stakeholders. Cross-Domain Sentiment Analysis (CSDA) is the process of analysing and interpreting sentiments in text data across different subject areas or contexts, accounting for the varying nuances and contextual differences in sentiment expression. The problem of CDSA poses a significant challenge in the field of Natural Language Processing (NLP), as the sentiment polarity of words and expressions can vary drastically across different domains. For instance, a word like "unpredictable" can convey positive sentiment in the context of a movie review but may signify negative sentiment when referring to the performance of a computer system. Deep Learning (DL), a subfield of machine learning, has shown promising results in various domains since its emergence in 2006, especially in complex problem-solving involving vast datasets. This paper aims to evaluate CDSA performance using Convolutional Neural Network (CNN) on the Amazon dataset. The study builds upon our previous research that highlighted the limitations of classical Machine Learning (ML) approaches for CDSA. The result demonstrates that the DL model is the state-of-the-art in machine learning classification tasks even though with a limited features engineering task. In conclusion, understanding people's opinions across different subjects on the internet is crucial but complex and using advanced Deep Learning methods like the Convolutional Neural Network can help address these challenges effectively.

1. Introduction

Sentiment analysis (SA) is the computational study of analysing people's feelings and opinions for an entity whereby the field of sentiment analysis has been the topic of extensive research in the past decades [1]. The process aims to understand a user's attitude and opinions by investigating, analysing and extracting subjective texts involving users' opinions, preferences and sentiment. It is a combination of the computational study of people's attitudes, appraisals and opinions about

* *Corresponding author*
E-mail address: azwaaziz@unisza.edu.my

https://doi.org/10.37934/ard.138.1.129136

individuals, issues, entities, topics, events, products and their attributes [2]. SA helps decision-makers in digesting this information, especially in real-time data monitoring. Pang *et al.,* [3], marked the beginning of widespread awareness of the research problems and opportunities that SA and opinion mining raise. With a tremendous number of online communities generating social media content constantly at high speed, understanding sentiments in social media content are valuable for customers, business owners and other stakeholders [4]. Campaign managers, politicians and even equity investors and online shoppers are the direct beneficiaries of SA technology.

Deep Learning (DL) is one of the fields that has emerged since 2006, where it produced breakthrough results in many domains [5]. It is a sub-field of machine learning algorithms exploring multiple layers of distributed representations. It is also known as hierarchical machine learning to solve a complex problem with the presence of big data and with very little human interaction. The DL performs by learning on multiple layers of nonlinear neural networks, which then will convert the form to a higher and more complex level. The learned form can be implemented as features and is useful in detection or classification activities. Additionally, deep neural networks are used to learn the input data's optimal expression, which can then be used to solve a complex problem [6].

DL has emerged as a powerful machine learning technique that learns multiple layers of representation or data features and produces a state-of-the-art result for sentiment classification. Since the performance of a machine learner heavily depends on the choices of data representation, many studies are devoted to building powerful feature extractors with a domain expert and careful engineering. DL discovers intricate semantic representations of texts automatically from data without feature engineering [7]. DL models such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RCN) have achieved state-of-the-art results for various natural language processing tasks including sentiment analysis. The result indicates highly accurate predictions are possible using these in conjunction with large datasets but with little understanding of the internal features and representations of the data that a model uses to classify into sentiment categories.

The main key challenge of sentiment analysis is that it is highly domain-dependent [2,8,9]. This relates to the ability of the SA model to perform very well in a specific domain (cross-domain sentiment), but very poorly in another or just mediocre in all domains. For example, the word *unpredictable* in "*the movie is great. The plot is unpredictable*" represents positive sentiment for the movie domain for the word *"unpredictable"*. In contrast, the same word defines negative sentiment in the computer domain such as "*The computer's performance is unpredictable*" which conveys that the machine performance is not consistent. This example demonstrates how the exact words can have different sentiment polarities for sentiment in other domains. There are usually different words and expressions in different domains and moreover, even the same word in different domains may reflect different sentiment polarities [10].

Based on work by Aziz *et al.,* [11], we evaluated cross-domain sentiment analysis using classical Machine Learning (ML); Random Forest Classifier (RFC), Multinomial Naïve Bayes (MNB), Stochastic Gradient Descent (SGD) and Support Vector Machine (SVM). The best result for cross-domain experiments only 70% by using SVM approaches.

Therefore, the paper focus implementing Deep Learning (DL) approaches specifically CNN using Amazon dataset to compare the results with classical ML approaches. CNN is chosen for this study based on the results achieved by Dang *et al.,* [12] in comparing DL models for sentiment analysis tasks. The result revealed that CNN outperforms other models, presenting a good balance between accuracy and CPU runtime.

2. Previous Works

The implementation of DL to solve the sentiment analysis problem is growing due to the fact that in recent years it is influential in both supervised and unsupervised learning domains [13]. Users' sentiment (or opinion) is always characterized by an unstructured form of data with an additional lack of labels for such data. Thus, integrating sentiment analysis with the DL algorithm seems promising due to its automatic learning capability. One of the early successful examples of implementation of DL with sentiment analysis, using RNN is in analysing movies reviews from the website rottentomatoes.com [14]. Figure 1 shows the comparison between classical NLP approaches for SA with DL.

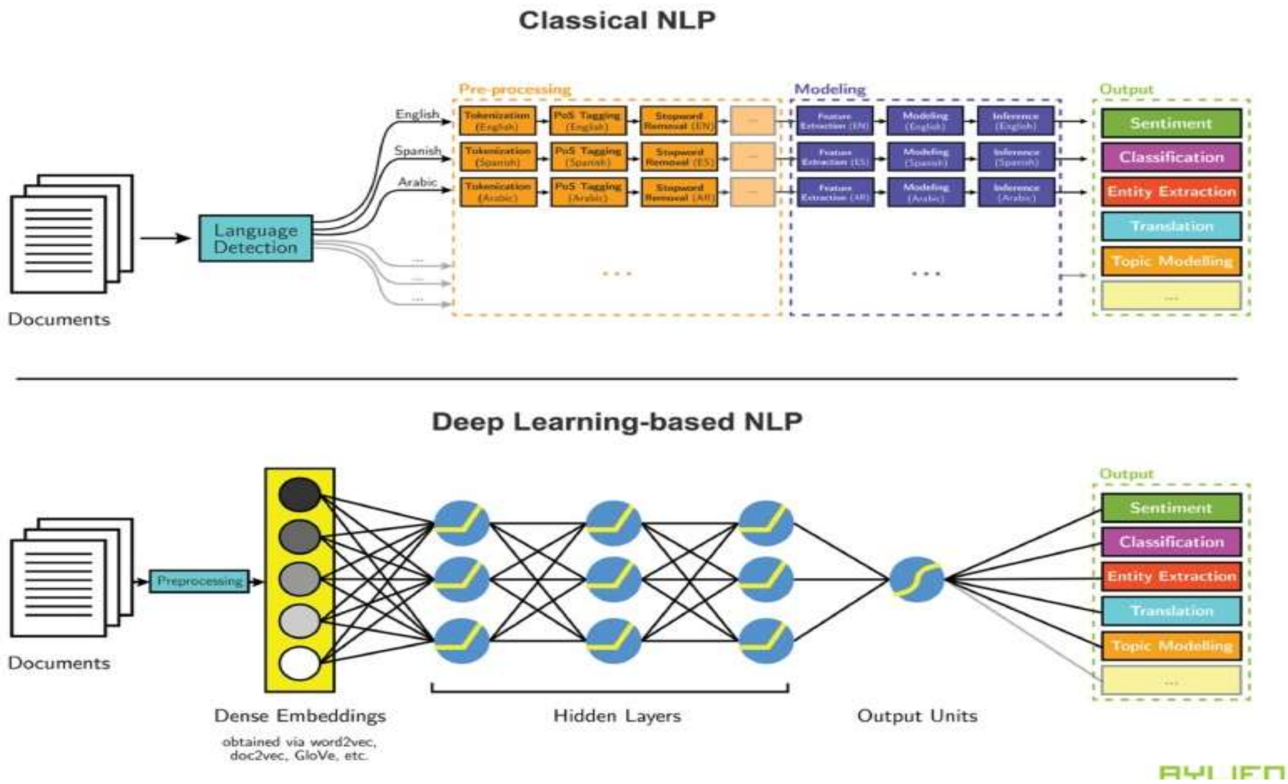

Fig. 1. Comparison between classical ML with DL for sentiment [15]

In classical NLP approaches, the critical process involves language detection, which requires substantial lexicon resources. The process continues by feature engineering processes such as Pos Tagging and exploring word relation (ex: word2Vec) to obtain higher sentiment accuracy. The classical NLP approaches have proven to be difficult, mainly due to the language-dependent nature of pre-processing and feature engineering techniques employed in traditional approaches. However, DL based NLP methods, which have gained a tremendous amount of growing attention and popularity over the last couple of years, have been proven to bring an incredible amount of invariance to NLP processes and pipelines, including towards the language used in a document or utterance [15].

DL learns multiple layers of representations or features of the data and produces state-of-the-art prediction results [7]. It is the part of the renaissance of the Artificial Neural Network (ANN) concept with high processing computing power and availability of a huge amount of training datasets that have contributed to more research focus on ANN methods. Neural Networks consist of a large number of information processing units (neurons) organised in layers to perform tasks such as classification.

In DL, the system can self-teach in that it can learn as it goes by filtering information through these multiple hidden layers. DL relies on the discovery that unsupervised learning could be used to set each level of a hierarchy of features, one level at a time, based on the features discovered at the previous level [16].

DL analysis capabilities with complex datasets have inspired researchers to apply it to sentiment analysis studies due to the high dimensionality features of text data. With the development of DL, many neural network-based sentiment analysis methods have achieved good results in public datasets; however, effective DL methods heavily depend on large labelled training data, which requires expensive manual labelling and is time-consuming. The approaches also generally require more computational resources and complexity of training dataset for supervised DL. Most sentiment analysis research using DL also needs to be combined with corpus approaches (Lexicon-Based) to identify the vectors.

3. Methodology

CNN is chosen for this study based on the results achieved by Dang *et al.,* [12] in comparing DL models for sentiment analysis tasks. The result revealed that CNN outperforms other models, presenting a good balance between accuracy and CPU runtime. Figure 2 shows the general process of CNN classification.

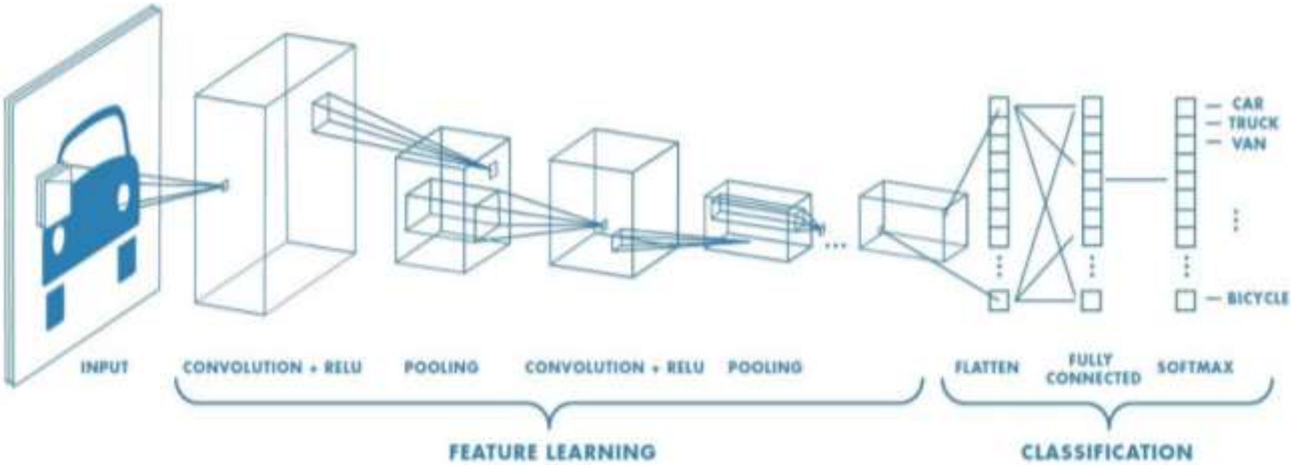

Fig. 2. CNN process [17]

CNN is a special type of feed-forward neural network employed initially in areas such as computer vision, recommender systems and natural language processing. It is a deep neural network architecture, typically composed of convolutional and pooling or subsampling layers to provide inputs to a fully connected classification layer [7]. Convolution layers filter their inputs to extract features; the outputs of multiple filters can be combined. Pooling or subsampling layers reduce the resolution of features, which can increase the CNN’s robustness to noise and distortion. Fully connected layers perform classification tasks. Two common pooling methods are average pooling and max pooling, which summarize a feature's average presence and the most activated presence of a feature, respectively. CNN does not require expert knowledge about a target language's linguistic structure, which becomes an advantage of using DL for sentiment classification [18].

The Rectified Linear Unit (ReLU) function is a non-linear activation function that has gained popularity in the DL domain. The main advantage of using the ReLU function over other activation functions such as sigmoid is that it does not activate all the neurons simultaneously. Thus, it can be more computationally efficient when compared to other functions. It means that the neurons will only be deactivated if the output of the linear transformation is less than 0. The neuron does not activate if there are negative values that are equal to the result is zero function as in Eq. (1).

$$f(x) = \max(0, x) \tag{1}$$

Another term that needs to be understood in the DL process is epoch. An epoch is used in machine learning to indicate the number of passes of the entire training dataset through the machine learning algorithm. Datasets are usually grouped into batches (especially when the amount of data is tremendous). Some people use the term iteration loosely and refer to putting one batch through the model as an iteration. If the batch size is the whole training dataset, then the number of epochs is the number of iterations.

The research carried out using Amazon cross-domain dataset. Amazon cross-domain data consists of four different domain review data (book, DVD, Electronic, Kitchen). The data originated from research by Blitzer *et al.,* [19]. There are 2000 positive and negative sentiment training data for each domain. Researchers widely use it in the sentiment analysis area, especially to measure classification for cross-domain data.

4. Result

Table 1 shows the cross-domain experiments using CNN models to find the best result for further analysis.

Table 1
Experiment result using CNN models

Model	Features Selection	Variables Setting	Layer	Epoch	Time	Accuracy(%)
CNN_1	Count vectorizer		ReLU	5	10s	70.2
CNN_2	Words Embedded	Embed dim = 100 Maxlen =100	ReLU	5	1m 50s	50.42
CNN_3	Words Embedded	Embed dim = 100 Maxlen =100	ReLU Maxpooling	5	1m 50s	74.22
CNN_4	Words Embedded	Embed dim = 100 Maxlen =100	ReLU Maxpooling Conv1d	5	2m 10s	70.12

Experiments show that using the CNN_3 model gives the highest result with embedding words as the feature selection method, ReLU as activation functions and Maxpooling as the convolution layers filter [20]. The 74.22% prediction accuracy already outperforms the highest result for SML or CA-HKT experiments, evidencing DL as the state-of-the-art in sentiment classification tasks. Therefore, CNN_3 is chosen to be analysed further by changing the settings and number of the variables of epochs to obtain the model with the highest accuracy.

Table 2 represents the results from 10 experiments of CNN_3 models using different variables and epochs settings. Experiment 10 gives the highest result with 77% (76.46%) accuracy using three epochs.

Journal of Advanced Research Design
Volume 138 Issue 1 (2026) 129-136

Table 2
Experiment result from CNN_3 model

Cnn_3 Model	Features Selection	Variables Setting	Layer	Epoch	Time	Accuracy (%)
1	Words Embedded	Embed dim = 100 Maxlen = 100	Relu Maxpooling	5	1m 50s	74.22
2	Words Embedded	Embed dim = 100 Maxlen = 150	Relu Maxpooling	5	2m 20s	73.76
3	Words Embedded	Embed dim = 150 Maxlen = 100	Relu Maxpooling	5	2m 20s	73.71
4	Words Embedded	Embed dim = 150 Maxlen = 150	Relu Maxpooling	5	2m 40s	74.99
5	Words Embedded	Embed dim = 200 Maxlen = 200	Relu Maxpooling	5	3m 40s	74.48
6	Words Embedded	Embed dim = 400 Maxlen = 400	Relu Maxpooling	5	5m 20s	74.68
7	Words Embedded	Embed dim = 150 Maxlen = 150	Relu Maxpooling	3	1m40s	75.6
8	Words Embedded	Embed dim = 250 Maxlen = 1150	Relu Maxpooling	3	3m10s	75.6
9	Words Embedded	Embed dim = 150 Maxlen = 150 Num_words=15000	Relu Maxpooling	3	1m40s	76.13
10	Words Embedded	Embed dim = 250 Maxlen = 1150	Relu Maxpooling	3	3m25s	76.46
11	Words Embedded	Embed dim = 150 Maxlen = 150	Relu Maxpooling	10	5m20s	73.23
12	Words Embedded	Embed dim = 150 Maxlen = 150	Relu Maxpooling	20	10m40s	72.9

CNN_3 model consistency gives results between 74 to 77%, with the highest being 76.46%. Comparing results with classical ML CNN_3 clearly surpasses all the models as shown in Figure 3.

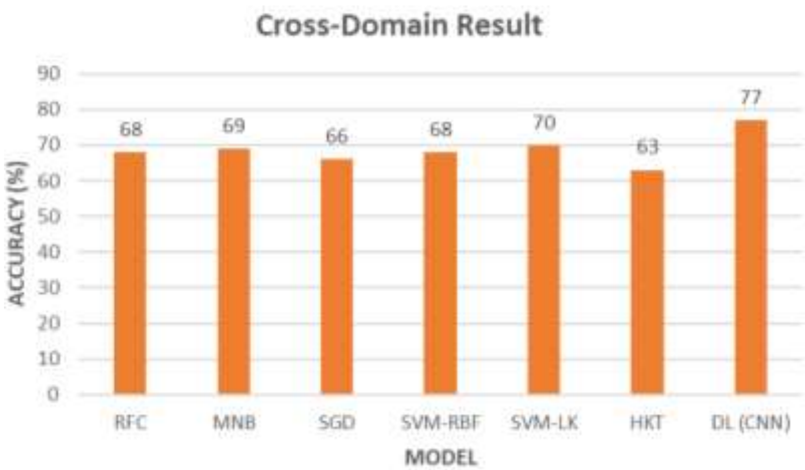

Fig. 3. Comparison between SML models

The result demonstrates that the DL model is the state-of-the-art in machine learning classification tasks although with limited features engineering. In other studies, the result for cross-domain experiments can achieve more than 90% with extra works in features engineering, including transfer learning or BERT as pre-training languages.

5. Conclusions

In conclusion, DL methods are proven as state-of-the art for cross-domain classification tasks. Additional feature engineering processes or integration with a pre-trained model such as BERT improves accuracy by more than 90%, as proven in previous studies. However, DL approaches suffer from a lack of transparency and are unable to explain how the result is achieved. In contrast, the abilities of CA-HKT to capture original sources and words during the classification process give the ability to reverse engineer any specific result in order to fully explain how the result was obtained.

In this journal, we have explored the application of Deep Learning, specifically CNN for sentiment using cross-domain datasets. The study began by highlighting the significance of SA in real-time data monitoring, decision-making and its wide-ranging applications across various domains, including business, politics and e-commerce. It was noted that the growth of online communities and social media has amplified the need for effective SA techniques. Deep Learning, with its ability to learn hierarchical representations of data, has emerged as a powerful approach for SA. CNNs have demonstrated their effectiveness in various natural language processing tasks, including sentiment analysis. While the internal features and representations learned by DL models may not always be interpretable, their high accuracy on sentiment classification tasks is evident.

The experiments with CNN models and various configurations, including feature selection, activation functions (ReLU) and convolution layers (Maxpooling), yielded promising results. The best-performing CNN model achieved an accuracy of 76.46%, outperforming classical ML approaches. Moreover, we observed that DL-based models, when combined with extensive feature engineering or advanced techniques like transfer learning and BERT, can achieve even higher accuracy, surpassing the 90% mark in cross-domain sentiment analysis.

However, the ‘black-box’ nature of Deep Neural Networks challenges its usefulness in mission-critical applications (ex: healthcare, self-driving machine, military), raising ethical and judicial concerns including lack of trust. Most techniques do not disclose how and why decisions are taken. In other words, these black-box algorithms lack transparency and explainability, leading to the development of the new emerging research known as Explainable Artificial Intelligence (XAI).

XAI demands human-understandable approaches for the machine tasks. XAI aims to address how AI systems decide which AI methods and techniques produce human-comprehensible solutions. XAI solutions will enable enhanced prediction accuracy with decision understanding and traceability of actions taken. XAI aims to improve human understanding, determine the justifiability of decisions made by the machine, introduce trust and reduce bias. The new machine-learning systems need to have the ability to explain their rationale, characterize their strengths and weaknesses and convey an understanding of how they will behave in the future. Therefore, for the future works, the exploration will focus on proposing Explainable DL model to provide transparent AI decision making model for sentiment.

Acknowledgement
This research was funded by a grant from Ministry of Higher Education of Malaysia Scheme [FRGS/1/2022/ICT02/UNISZA/02/2].

References
[1] Ligthart, Alexander, Cagatay Catal and Bedir Tekinerdogan. "Systematic reviews in sentiment analysis: a tertiary study." *Artificial intelligence review* (2021): 1-57.

[2] Al-Moslmi, Tareq, Nazlia Omar, Salwani Abdullah and Mohammed Albared. "Approaches to cross-domain sentiment analysis: A systematic literature review." *Ieee access* 5 (2017): 16173-16192. https://doi.org/10.1109/ACCESS.2017.2690342

[3] Pang, Bo and Lillian Lee. "Opinion mining and sentiment analysis." *Foundations and Trends® in information retrieval* 2, no. 1–2 (2008): 1-135. https://doi.org/10.1561/1500000011

[4] Tao, Jie and Xing Fang. "Toward multi-label sentiment analysis: a transfer learning based approach." *Journal of Big Data* 7, no. 1 (2020): 1. https://doi.org/10.1186/s40537-019-0278-0

[5] Day, Min-Yuh and Chia-Chou Lee. "Deep learning for financial sentiment analysis on finance news providers." In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1127-1134. IEEE, 2016. https://doi.org/10.1109/ASONAM.2016.7752381

[6] Shirahatti, Abhinandan and Krupa Rasane. "A survey of deep learning for sentiment analysis." *Asian Journal For Convergence In Technology (AJCT) ISSN-2350-1146* (2019).

[7] Zhang, Lei, Shuai Wang and Bing Liu. "Deep learning for sentiment analysis: A survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, no. 4 (2018): e1253. https://doi.org/10.1002/widm.1253

[8] Save, Ashwini and Narendra Shekokar. "Analysis of cross domain sentiment techniques." In *2017 international conference on electrical, electronics, communication, computer and optimization techniques (ICEECCOT)*, pp. 1-9. IEEE, 2017. https://doi.org/10.1109/ICEECCOT.2017.8284637

[9] Yuan, Zhigang, Sixing Wu, Fangzhao Wu, Junxin Liu and Yongfeng Huang. "Domain attention model for multi-domain sentiment classification." *Knowledge-Based Systems* 155 (2018): 1-10. https://doi.org/10.1016/j.knosys.2018.05.004

[10] Kadli, Parvati and B. M. Vidyavathi. "Deep-Learned Cross-Domain Sentiment Classification Using Integrated Polarity Score Pattern Embedding on Tri Model Attention Network, vol. 12." (2021): 1910-1924. https://doi.org/10.21817/indjcse/2021/v12i6/211206190

[11] Aziz, Azwa Abdul andrew Starkey and Marcus Campbell Bannerman. "Evaluating cross domain sentiment analysis using supervised machine learning techniques." In *2017 Intelligent Systems Conference (IntelliSys)*, pp. 689-696. IEEE, 2017. https://doi.org/10.1109/IntelliSys.2017.8324369

[12] Dang, Nhan Cach, María N. Moreno-García and Fernando De la Prieta. "Sentiment analysis based on deep learning: A comparative study." *Electronics* 9, no. 3 (2020): 483. https://doi.org/10.3390/electronics9030483

[13] Ain, Qurat Tul, Mubashir Ali, Amna Riaz, Amna Noureen, Muhammad Kamran, Babar Hayat and A. Rehman. "Sentiment analysis using deep learning techniques: a review." *International Journal of Advanced Computer Science and Applications* 8, no. 6 (2017). https://doi.org/10.14569/IJACSA.2017.080657

[14] Socher, Richard, Cliff C. Lin, Chris Manning and Andrew Y. Ng. "Parsing natural scenes and natural language with recursive neural networks." In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 129-136. 2011.

[15] Aylien.com. "Leveraging Deep Learning for Multilingual Sentiment Analysis." (2020). https://aylien.com/blog/leveraging-deep-learning-for-multilingual

[16] Glorot, Xavier, Antoine Bordes and Yoshua Bengio. "Domain adaptation for large-scale sentiment classification: A deep learning approach." In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 513-520. 2011.

[17] Saha, Sumit. "A comprehensive guide to convolutional neural networks—the ELI5 way." *Towards data science* 15 (2018): 15.

[18] Kim, Hannah and Young-Seob Jeong. "Sentiment classification using convolutional neural networks." *Applied Sciences* 9, no. 11 (2019): 2347. https://doi.org/10.3390/app9112347

[19] Blitzer, John, Mark Dredze and Fernando Pereira. "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification." In *Proceedings of the 45th annual meeting of the association of computational linguistics*, pp. 440-447. 2007.

[20] Khandenwal, R. "Words embedding for NLP, toward data science." (2019). https://towardsdatascience.com/word-embeddings-for-nlp-5b72991e01d4