

Efficient Deep Learning Models for Extreme Multi-Label Text Classification - Review Paper

Shumama Ansa¹ and Dr. G. Narsimha²

¹Research Scholar, Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad, Telangana, India

Email: shumama2011@gmail.com

²Professor and Principal, Computer Science and Engineering, JNTUH University College of Engineering Sultanpur, Telangana, India

Abstract— While social media has enabled the broad dissemination of user perspectives, it has also generated an immense volume of unstructured text requiring sophisticated categorization techniques. This comprehensive literature review explores advances in deep learning models tailored for extreme multi-label text classification of social media content. Significant progress includes the development of neural architectures like convolutional networks, re-current networks, and transformer models such as BERT, markedly improving the scalability, accuracy, and interpretability of classification tasks. Transformer models like BERT have transformed the field by facilitating a richer comprehension of multifaceted social media posts, enabling more precise label predictions. However, classifying social media presents unique difficulties due to its high dimensionality, and sparse and imbalanced distributions that challenge model training. Emerging approaches address these hurdles, such as label partitioning, attention mechanisms, and hybrid architectures, strengthening model robustness and efficiency. Nonetheless, building interpretable, real-time applicable systems capable of adapting to social media's dynamic nature remains an ongoing challenge. This analysis provides a thorough overview of the deep learning state-of-the-art for multi-label classification and outlines future directions to resolve lingering issues in this rapidly advancing domain. The findings are intended to guide next-generation model development better suited to social media's intricate classification demands.

Keywords— Extreme Multi-Label Text Classification (XMTC), Multi-Label Classification, Deep Learning, Transformer-based Architectures, Transfer Learning.

I. INTRODUCTION

With the onset of the digital information age, social media has proved to be a powerful instrument of communication characterized by vast amounts of user-generated data. The recent rise of social media content has also opened up exciting challenges and opportunities for information extraction and analysis. One of the key processes that involved with this data is to efficiently recognize and categorize text, particularly more so in XMTC. XMTC describes when hundreds of thousands or even millions of possible categories, the method for assigning a set of relevant labels to each document from an enormous universe of labels [1]. As a result, this presents unique challenges, in the form of data sparsity and scalability problems, that have driven advances in innovative machine learning techniques.

Text classification is an important task that goes beyond mere categorization. It has wide-ranging implications on multiple application domains, such as information retrieval, question answering, text summarization, and natural language understanding [3].

In the social media world, accurate categorization can play key roles in different areas like business analytics, healthcare surveillance and public safety [4].

Over the few past years a shift to deep learning methodologies for complex text categorization tasks. CNN, LSTM and their variations showed encouraging results in handling the nuances of social media content [5]. These models are able to recognize complex patterns and relationships in data usually out-smarting traditional machine learning methods. Deep Learning is of high importance in the field XMTc Over various benchmark datasets, CNN-based methodologies have always produced state-of-the-art results. This example showcases the power of deep learning in addressing challenges posed by extreme multi-label scenarios.

However, their architectural configuration is not the only factor that will determine how effective these models are. It is where text representation and feature extraction techniques play the key part. We study methods including Term Frequency-Inverse Document Frequency (TF-IDF), Word Article Matrix (WAM) and word embedding models such as Word2Vec, GloVe, and fastText to better classify them [6].

At the same time, bringing more contextual information in advance of simple text improved classification results. Deep learning frameworks integrated with user behavioral data and sentiment analysis have been proposed to improve the interpretation of social media messages [7]. In practice, several label sets may include a hierarchical structure, which adds further complexity to the assignment of categories. Methods such as Hierarchical Label Set Expansion (HLSE) have been proposed to improve label refinement and increase classification performance in these settings [8].

Challenges: Despite the significant recent progress made in the field of extreme multi-label text categorisation, challenges still remain.

- Need of richer, enhanced method of training with larger datasets
- Models that generalize across multiple domains and languages
- Ability to deal with changes in label spaces over time (e.g., in liquid social media settings)

In this literature review, we aim to give a complete overview on the recent progress in terms of deep learning architectures for extreme multi-label classification on social media content detection. In this post, we will review many different architectures of models, and methods for feature extraction and evaluation both in past research literature.

In addition, we will evaluate the difficulties and limitations of current procedures and point out potential new directions for further studies. This study aims to detail the most relevant approaches for dealing with XMTc in social media contexts integrating information from several research works. We will test our approach on various applications — besides broad topic categorization, we selected more specialized tasks, like extremist material detection [9] or identifying violent posts [10].

Furthermore, we will investigate the potential of novel techniques such as:

- Cross-event Twitter topic classification using CNN models [11]
- Feature-reduction approaches for violence detection in Arabic dialects [12]
- Bidirectional association-based pooling layers for acquiring high-level features and labels [13]

These techniques aim to improve classification performance without relying on explicit topological organization of the data. Through this comprehensive review, we aim to provide researchers and practitioners with a solid foundation for developing and implementing high-efficiency DL architectures for intense multi-label social media text classification. Our findings will contribute to the ongoing efforts to harness the extensive data provided in social media, ultimately leading to more effective and insightful analysis of online discourse.

II. BACKGROUND AND RELATED WORKS

The advancement of social media has led to a remarkable amount of user-produced information, creating both possibilities and difficulties in text classification. Extreme Multi-Label Text Classification (XMTc) has become crucial in managing and analyzing this extensive unstructured data. [14] outlines various processes for sentiment analysis using popular machine learning algorithms on social media content. The study compares existing ML algorithms in Assessing digital reviews, finding that SVM outperforms Naive Bayes in accuracy.

In a related advancement, [15] introduces a news text classification method that combines deep learning algorithms. This study presents a customized DCLSTM-MLP model that processes text with both word vector and word dispersion. This approach addresses issues related to text length and feature extraction, achieving improved accuracy, recall, and overall performance compared to other models. [16] focuses on detecting and

classifying toxicity in social media using Bidirectional Encoder Representations from Transformers (BERT). This research fine-tunes the BERT and its weights in a toxic comment labeled dataset, evaluates the models to detect and evaluate the toxicity of twitter data surrounding UK Brexit incorporating how sensitive model can be used for leveraging detrimental effects of social media discourse due to toxicities.

[17] introduces the DeepXML architecture to deal with scalability and accuracy issues in deep extreme multi-label learning. We argue that this paradigm breaks down the problem into four pieces, leading to algorithms that trade accuracy for scalability (in several ways). The Astec algorithm designed within the framework observed large improvements in accuracy and speed of training over standard deep extreme classifiers, which suggest that it can be potentially useful in industrial-level implementations like Bing.

A methodology for sentiment classification on twitter tweets utilizing natural language processing and deep learning introduced[18]. The study employs distant supervision with Twitter messages as training data and achieves approximately 80% accuracy using the LSTM algorithm, showcasing the potential of leveraging readily available data for effective sentiment analysis in microblogging platforms. [19] re-views sentiment analysis of social media data for detecting apprehensiveness or dejection using di-verse AI methods, DL algorithms yields superior outcomes, underscoring effectiveness of advanced techniques in nuanced sentiment detection.

[20] explores multi-class sentiment classification using a multi-layer classification model. The study implements residing methodologies for sentiment analysis of movie reviews, showing that the multi-tier model slightly improves over single-layer architectures and offers better recall, enhancing the model's contextual learning. [21] uses transfer learning models, including various DL frameworks for sentiment analysis based on images from Twitter. The fine-tuned DenseNet-121 model outperforms others in im-age sentiment identification, achieving an accuracy of 0.89.

Additionally, [22] results indicate that classic machine learning models can match or even surpass neural networks in certain contexts.

Social media applications of the Seeker Optimisation Algorithm with Sentiment Analysis and Classification (SO-SAC) are recently presented in [23]. For model, it use a Self-Head Multi-Attention based Gated Recurrent Unit (SHMA-GRU) for identification and categorization of sentiment and uses Seeker Optimization Algorithm to optimize the hyperparameters in order to improve efficiency of classifier. [24] develops an SVM approach for fake news detection using a dataset of news texts and labels, comparing K-Nearest Neighbor and SVM methods and focusing on both news titles and text for a more comprehensive fake news detection system.

Further, [25] introduces an active learning-based pipeline to address data scarcity for new labels and utilizes ChatGPT for model evaluation, demonstrating the effectiveness of these techniques in enhancing classification tasks. [26] suggests a supervised deep learning classifier combining CNN and LSTM for multi-class sentiment analysis attains 85.8% accuracy and an F1 score of 0.86 on a labeled dataset of 42,036 Facebook comments, outperforming baseline models and demonstrating the proposed architecture's effectiveness for sentiment analysis in low-resource languages.

[27] Presents 4 deep learning techniques to perform sentiment analysis SwiftFFN indicates that it can do ordered statistical tests on more BERT variants with larger improvements in bigger datasets than CAG10k, while BiGRU achieves increased accuracy over both pre-trained BERT models and baselines of traditional machine learning approaches. The effectiveness of the most common text classification algorithms is examined by [28], in social theory research domain where little labeled data and many low-frequency categories are available and results show that, supervised ML approaches are better than state-of-art lexicons – for deep learning architectures which their complexity is high as compare to logistic regression or random forest baselines (which demand less computational effort) perform equally well.

A hybrid deep learning network model named “Fasttext Convolutional Neural Network with Long Short-Term Memory (FCL)” for depression identification using textual data design proposed in [29]. A similar work in [30] shows a transformer-based solution for multi-labelisation of offensive texts in Bengali. To evaluate ours and other computational strategies, we introduce a dataset called EM-BAD containing 13,728 texts labeled in five target categories. With a refined pooling layer that actually allows for fine-tuning, Bangla-BERT beats every baseline (ML+ DL + Transformer) model. [31] for the impact of machine learning classifiers, as well as deep learning architectures. The deep learning models alms to be realised in the study range from Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. Machine learning and deep learning models will be developed to evaluate using a Covid19 dataset containing 1,375,592 tweets.

Table 1 provides the summary of distinct existing techniques handled by the researchers. The area of extreme multi-label text categorization for social media is rapidly evolving, with deep learning architectures driving progress. Recent progress in extreme multi-label text categorization for social media involves fast-moving frontiers, largely driven by innovations in deep learning architectures. With astonishing development in deep

learning architectures, the domain of extreme multi-label text categorization for social media is swiftly advancing. As we continue to grapple with the issues caused by the scale and complexity of social media data, enhancing more efficient, accurate, and adaptable classification models is a critical area of research.

TABLE I: SURVEY FROM DIFFERENT AUTHORS

S.No	Reference	Methodology	Results Obtained	Limitations	year
1	[15]	Combination of CNN, LSTM, and MLP	Effective Multi-Label Classification with an accuracy of 94.28	High computational requirements	2021
2	[16]	BERT and its variants	Private score of 98.6 Efficient Computational Resource Utilization	Potential bias in toxicity definition	2021
3	[17]	Deep XML framework	2-12% higher accuracy and 5 to 30 times greater speed compared to leading deep extreme classifiers	Scalability with extremely large label sets	2021
4	[20]	Decision Tree, SVM, Naïve Bayes	Slight improvement over single-layer architecture, estimated 80-85% accuracy	Complexity in multi-tier architecture	2022
5	[21]	VGG-19, ResNet50V2, DenseNet-121	DenseNet-121: 0.89 accuracy, VGG-19: 0.73, ResNet50V2: 0.75	Challenges with multimodal data	2022
6	[22]	Classic machine learning vs deep learning	Classic ML on par or outperforming neural networks, estimated 75-80% accuracy	Limited to under-resourced languages	2022
7	[23]	SHMA-GRU model with SOA	SOADL-SAC model demonstrated an accuracy of 79.46% on sentiment 140 dataset.	Complexity of optimization process	2022
8	[24]	SVM approach	Enhanced Interpretability	Challenges with fake news detection	2023
9	[25]	Label ranking model, active learning, ChatGPT for evaluation	Scalability for Large Datasets. performance boost of approximately 15 points, resulting in a Recall@10 of 0.85.	Greedy acquisition strategy, inefficient use of negative samples	2023
10	[26]	CNN and LSTM	Achieved 85.8% accuracy, 0.86 F1 score	Limited to Bengali language	2023
11	[27]	BERT with BiLSTM and BiGRU	Of the four DistilBERT models, GLG, has 88.04% accuracy. Of the four RoBERTa models, 3G, has 91.72% accuracy	High computational requirements	2023
12	[29]	FCL (Fasttext CNN LSTM)	FCL model attained 87% accuracy.	Ethical concerns with mental health detection	2024

III. SYSTEMATIC LITERATURE REVIEW

SLR is a crucial process for synthesizing existing research, providing a comprehensive overview of advancements, challenges, and future directions within a specific field. This SLR focuses on the development of DL architectures for XMTC in social media contexts. The review aims to evaluate the current state of research, identify key challenges, and explore potential opportunities for enhancing the efficiency and accuracy of XMTC models.

A. Overview of Selected Studies

The selected studies cover various aspects of deep learning architectures and their application to extreme multi-label text classification in social media contexts. Key observations include:

- **Advancements in Deep Learning Architectures:** The review highlights significant progress in the development of deep learning models, such as CNNs, RNNs, and particularly Transformer-based models like BERT, exhibited effective performance in handling large-scale, complex label sets.
- **Transformer-based Models and BERT:** Transformer models, including BERT, have revolutionized XMTC by enabling better contextual understanding of text, which is crucial for accurately classifying highly dynamic and diverse social media content.
- **Challenges and Emerging Trends:** Challenges such as data sparsity, label imbalance, and scalability are prominent in XMTC. However, emerging trends, such as attention mechanisms and hybrid models, offer promising solutions to these issues.

B. Comparative Analysis

Benchmarking methodologies, accuracy metrics, and scalability of the selected studies reveals:

- Scalability: While Transformer-based models improve accuracy, scalability remains a challenge, particularly for real-time applications on large datasets. Optimization techniques and lightweight models are being explored to address these issues.
- Accuracy: BERT and other Transformer-based models demonstrate high accuracy in XMTC tasks, outperforming traditional models. However, their effectiveness varies depending on the dataset and the complexity of the label set.
- Interpretability: The complexity of deep learning models, especially Transformers, poses challenges for interpretability, which is crucial for real-world applications.

IV. FINDINGS

Data extraction was performed on the selected research articles, and the results are discussed. Also, a gap analysis is provided to highlight areas requiring further research and development.

A. Gap Analysis

Despite significant advancements in XMTC for social media, several gaps remain in the current research landscape. There is a need for more efficient and adaptable models that can handle the rapidly evolving nature of social media content and label spaces. The trade-off between model complexity and computational efficiency remains a challenge, particularly for real-time applications. Interpretability of deep learning models in XMTC tasks is still limited, hindering their adoption in sensitive domains. Additionally, there is a lack of standardized benchmarks specifically designed for social media XMTC tasks, making it difficult to compare different approaches consistently. Future research should focus on developing more robust and efficient architectures that can balance accuracy, scalability, and interpretability while addressing the unique challenges posed by social media data.

B. Future Work

- Computational Aware Deep Learning Model can be designed for solving the high dimensional big data vectors.
- Efficient Ensemble XML Deep learning model can be designed to remove the bottleneck of deployment computational cost and to achieve efficient equal distribution of labels.

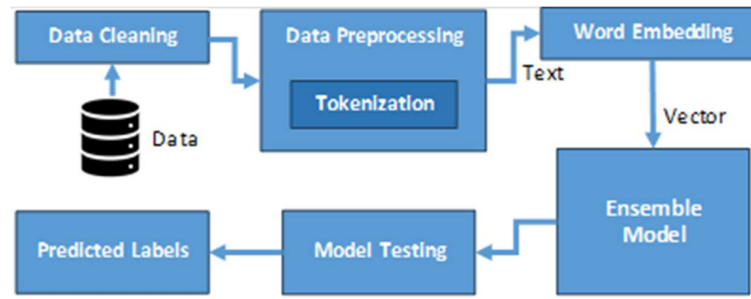


Figure 1: Procedure for Extreme multi label classification using Ensemble model.

We can consider a dataset named Goemotions which can be downloaded from kaggle website. The Google AI GoEmotions dataset consists of comments from Reddit users. The categories of emotions include 12 positive, 11 negative, 4 ambiguous emotions, and 1 neutral, which makes the dataset suitable for solving tasks that require subtle differentiation between different emotions. Ensemble deep learning model can be applied to the dataset to achieve efficient equal distribution of labels.

V. CONCLUSION

This systematic literature review explored the development of high-efficient deep learning architectures for extreme multi-label social media text classification. The analysis identified significant advancements in Transformer-based models, particularly BERT, and their impact on improving classification accuracy. Despite these advancements, challenges such as scalability, data sparsity, and interpretability persist. Future research

should focus on optimizing Transformer-based models for large-scale datasets, improving the interpretability of deep learning architectures, and developing scalable solutions to meet the demands of real-time social media text classification. Additionally, hybrid models and transfer learning approaches show promise in addressing these challenges. The review underscores the potential of these advanced architectures to transform social media text classification while highlighting the need for ongoing innovation.

REFERENCES

- [1] Liu Jingzhou, Chang Wei-Cheng, Wu Yuexin, Yang Yiming. Deep Learning for Extreme Multi-label Text Classification. 115-124. 10.1145/3077136.3080834, 2017
- [2] Jotikabukkana, Phat & Sornlertlamvanich, Virach & Manabu, Okumura & Haruechaiyasak, Choochart. "Social Media Text Classification by Enhancing Well-Formed Text Trained Model". Journal of ICT Research and Applications. 10. 177-196. 10.5614/itbj.ict.res.appl.2016.10.2.6, 2016
- [3] Pérez-Vera, S., Alfaro, R., Allende-Cid, H. "Intent Classification of Social Media Texts with Machine Learning for Customer Service Improvement". In: Meiselwitz, G. (eds) Social Computing and Social Media. Applications and Analytics. SCSM 2017. Lecture Notes in Computer Science(), vol 10283. Springer, Cham. https://doi.org/10.1007/978-3-319-58562-8_21, 2017.
- [4] T.Y.S.Santosh Vaibhav Gambhir Animesh Mukherjee, "Deep Learning for social media health text classification" Proceedings of the 3rd Social Media Mining for Health Applications (SMM4H) Workshop & Shared Task, pages 61–64 Brussels, Belgium, October 31, 2018.
- [5] Tsapatoulis N, Djouvas C, "Opinion Mining From Social Media Short Texts: Does Collective Intelligence Beat Deep Learning?" Front Robot AI. 2019 Jan 22;5:138. doi: 10.3389/frobt.2018.00138. PMID: 33501016; PMCID: PMC7805642., 2019.
- [6] Ahmed Sulaiman M. Alharbi, "Social Media Sentiment Analysis with a Deep Neural Network: An Enhanced Approach Using User Behavioral Information", Ph.D. Dissertation. Western Michigan University, USA. Order Number: AAI28108310, 2019.
- [7] Francesco Gargiulo, Stefano Silvestri, Mario Ciampi, Giuseppe De Pietro, "Deep neural network for hierarchical extreme multi-label text classification", Applied Soft Computing, Volume 79, Pages 125-138, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2019.03.041>, 2019
- [8] Ahmad, S., Asghar, M.Z., Alotaibi, F.M. et al., "Detection and classification of social media-based extremist affiliations using sentiment analysis techniques", Hum. Cent. Comput. Inf. Sci. 9, 24, <https://doi.org/10.1186/s13673-019-0185-6>, 2019.
- [9] Manzhu Yua , Qunying Huangb , Han Qina,c, Chris Scheeleb and Chaowei Yang, "Deep learning for real-time social media text classification for situation awareness – using Hurricanes Sandy, Harvey, and Irma as case studies." International journal of digital earth. vol. 12, no. 11, 1230–1247 <https://doi.org/10.1080/17538947.2019.1574316>, 2019
- [10] AL-Saif, H., & Alotaibi, T, "Arabic Text Classification using Feature-Reduction Techniques for Detecting Violence on Social Media". International Journal of Advanced Computer Science and Applications, 2019.
- [11] Ms. Mrunali D. Chaure, Prof. Jayant P. Mehare. "Text Classification and Analysis with Social Media Platform". IJRAR, Volume 6, Issue 4, December 2019.
- [12] Eduan KotzéI; Burgert A. SenekalII; Walter Daelemans, "Automatic classification of social media reports on violent incidents in South Africa using machine learning" South African Journal of Science S. Afr. j. sci. vol.116 n.3-4 Pretoria, <http://dx.doi.org/10.17159/sajs.2020/6557>, 2020.
- [13] Marilyn Bello, Gonzalo Nápoles, Ricardo Sánchez, Rafael Bello, Koen Vanhoof, "Deep neural network to extract high-level features and labels in multi-label classification problems", Neurocomputing, Volume 413, Pages 259-270, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2020.06.117>, 2020.
- [14] Bhardwaj, Aditya, "Sentiment Analysis and Text Classification for Social Media Contents Using Machine Learning Techniques", Proceedings of the 2nd International Conference on IoT, Social, Mobile, Analytics & Cloud in Computational Vision & Bio-Engineering (ISMAC-CVB 2020), Available at SSRN: <https://ssrn.com/abstract=3735851> or <http://dx.doi.org/10.2139/ssrn.3735851>, November 23, 2020.
- [15] Zhang, Menghan, "Applications of Deep Learning in News Text Classification", Scientific Program-ming, 6095354, 9 pages, 2021. <https://doi.org/10.1155/2021/6095354>, 2021.
- [16] Fan, H.; Du, W.; Dahou, A.; Ewees, A.A.; Yousri, D.; Elaziz, M.A.; Elsheikh, A.H.; Abualigah, L.; Al-qaness, M.A.A, "Social Media Toxicity Classification Using Deep Learning: Real-World Application UK Brexit". Electronics 2021, 10, 1332. <https://doi.org/10.3390/electronics10111332>, 2021.
- [17] Kunal Dahiya, Deepak Saini, Anshul Mittal, Ankush Shaw, Kushal Dave, Akshay Soni, Himanshu Jain, Sumeet Agarwal, and Manik Varma, "DeepXML: A Deep Extreme Multi-Label Learning Framework Applied to Short Text Documents" In Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21). Association for Computing Machinery, New York, NY, USA, 31–39. <https://doi.org/10.1145/3437963.3441810>, 2021.
- [18] Santhoshini , A. Shirisha , T. Sai meghana, "TEXT CLASSIFICATION USING DEEP LEARNING.", International Journal for Innovative Engineering & Management Research. Vol11 Issue 06, ISSN 2456 – 5083, 2022.

- [19] Babu, N.V., Kanaga, E.G.M, "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review", SN COMPUT. SCI. 3, 74, <https://doi.org/10.1007/s42979-021-00958-1>, 2022
- [20] Rahman, Hameedur & Tariq, Junaid & Masood, Ali & Subahi, Ahmad & Khalaf, Osamah & Alotaibi, Youseef, "Multi-Tier Sentiment Analysis of Social Media Text Using Supervised Machine Learning". Computers, Materials and Continua. 74. 5527-5543. 10.32604/cmc.2023.033190, 2022
- [21] Chandrasekaran, G.; Antoanela, N.; Andrei, G.; Monica, C.; Hemanth, J, "Visual Sentiment Analysis Using Deep Learning Models with Social Media Data", Appl. Sci. 2022, 12, 1030. <https://doi.org/10.3390/app12031030>, 2022.
- [22] Yantseva, V.; Kucher, K. Stance, "Classification of Social Media Texts for Under-Resourced Scenarios in Social Sciences". Data 2022, 7, 159. <https://doi.org/10.3390/data7110159>, 2022.
- [23] Hanan M. Alghamdi1, Saadia H.A. Hamza2, Aisha M. Mashraqi3 and Sayed Abdel-Khalek, "Seeker Optimization with Deep Learning Enabled Sentiment Analysis on Social Media", Computers, Materials & Continua DOI:10.32604/cmc.2022.031732, 2022.
- [24] M. Jagadeesan, T. M. Saravanan, P. A. Selvaraj, S. Nithya, K. Snegha and P. Sowmiya, "Social Media News Classification Using Machine Learning Algorithms", International Conference on Computer Communication and Informatics (ICCCI), pp. 1-5, doi: 10.1109/ICCCI56745.2023.10128245, 2023.
- [25] Dan Li1 ,Zi Long Zhu , Janneke van de Loo, Agn s Masip G mez, Vikrant Yadav, Georgios Tsatsaronis, Zubair Afzal, "Enhancing Extreme Multi-Label Text Classification: Addressing Challenges in Model, Data, and Evaluation", Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track, pages 313–321 December 6-10, 2023.
- [26] Rezaul Haque, Naimul Islam, Mayisha Tasneem, Amit Kumar Das, "Multi-class sentiment classification on Bengali social media comments using machine learning", International Journal of Cognitive Computing in Engineering, Volume 4, Pages 21-35, ISSN 2666-3074, <https://doi.org/10.1016/j.ijcce.2023.01.001>, 2023.
- [27] Talaat, A.S, "Sentiment analysis classification system using hybrid BERT models" J Big Data 10, 110. <https://doi.org/10.1186/s40537-023-00781-w>, 2023.
- [28] Shyrokikh K, Gimyk M, Dellmuth L, "Short text classification with machine learning in the social sciences: The case of climate change on Twitter" PLoS One. 29;18(9):e0290762. doi: 10.1371/journal.pone.0290762. PMID: 37773969; PMCID: PMC10540966, 2023.
- [29] Vankayala Tejaswini, Korra Sathya Babu, and Bibhudatta Sahoo, "Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model", ACM Trans. Asian Low-Resour. Lang. Inf. Process. 23, 1, Article 4, 20 pages. <https://doi.org/10.1145/3569580>, 2024.
- [30] Hossain, J., Das, A., Hoque, M. M., & Siddique, N. " Multilabel Aggressive Text Classification from Social Media using Transformer-based Approaches", In 2023 26th International Conference on Computer and Information Technology (ICCIT) (pp. 1-6). (2023 26th International Conference on Computer and Information Technology (ICCIT)). IEEE Advance online publication. <https://doi.org/10.1109/iccit60459.2023.10441608>, 2023.
- [31] M. Sudhakar, K.P. Kaliyamurthi, "Detection of fake news from social media using support vector machine learning algorithms", Measurement: Sensors, Volume 32, 101028, ISSN 2665-9174, <https://doi.org/10.1016/j.measen.2024.101028>, 2024.
- [32] Wang, Zhiqiang & Pang, Yiran & Lin, Yanbin, "Smart Expert System: Large Language Models as Text Classifiers", 2024.