JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Advanced Deep Learning Approaches for Sentiment Analysis in Code-Mixed Hinglish Text: A Comparative Study

Adarsh Singh Jadon
Computer Science & Engineering
Madhav Institute of Technology & Science
Prof. Mahesh Parmar
Computer Science & Engineering
Madhav Institute of Technology & Science
Dr. Rohit Agrawal
Computer Science & Engineering
Madhav Institute of Technology & Science

Abstract

This study investigated the efficacy of various deep learning models in performing sentiment analysis on code-mixed Hinglish text, a hybrid language widely used in digital communication. Hinglish presents unique challenges due to its informal nature, frequent code-switching, and complex linguistic structure. This research leverages datasets from the SemEval-2020 Task 9 competition and employs models such as RNN (LSTM), BERT-LSTM, CNN, and a proposed Hybrid LSTM-GRU model. The study's primary objective is to developed a robust sentiment analysis framework that accurately classifies sentiment in Hinglish text. The Hybrid LSTM-GRU model, combining the strengths of LSTM and GRU units, demonstrated superior performance with an accuracy of 92%, precision of 92.29%, and recall of 92.31%. This model outperformed existing approaches, including the HF-CSA model from the SemEval-2020 dataset, which achieved an accuracy of 76.18%. The results highlight the proposed model's capability to handle the nuances of Hinglish, including its informal and code-mixed nature, more effectively than traditional models. This study sets a new benchmark for sentiment analysis in multilingual contexts and underscores the importance of advanced deep learning techniques in tackling the challenges of code-mixed language processing.

Keywords: Hinglish, Sentiment Analysis, Deep Learning, Code-Mixed Language, RNN, LSTM, GRU, BERT, Neural Networks, Multilingual NLP, Informal Language Processing.

1.Introduction

In recent years, the proliferation of digital communication, particularly on social media platforms, has led to an increase in the use of code-mixed languages. One such language is Hinglish, a blend of Hindi and English, widely spoken in India and other South Asian countries. Hinglish combines elements from both languages, often within the same sentence, creating unique linguistic constructs that pose significant challenges for natural language processing (NLP) tasks, including sentiment analysis. The informal nature and frequent code-switching in Hinglish add layers of complexity, making it difficult for conventional sentiment analysis models to accurately interpret and classify the emotional tone of such texts.

Previous studies have explored various approaches to sentiment analysis in code-mixed languages. For instance, the HF-CSA model, developed during the SemEval-2020 Task 9 competition, utilized lexicon-based methods combined with machine learning techniques to handle code-mixed data (Raza et al., 2023). However, this model achieved only moderate accuracy, highlighting the need for more sophisticated approaches to tackle the intricacies of informal and cross-lingual text.

In response to these challenges, recent advancements in deep learning have introduced powerful models capable of capturing complex patterns in data. Recurrent Neural Networks (RNNs), particularly those with Long Short-Term Memory (LSTM) units, have shown promise in handling sequential data by effectively capturing long-term dependencies (Shrestha & Nasoz, 2019). Additionally, models that integrate Bidirectional Encoder Representations from Transformers (BERT) with LSTM have demonstrated the ability to leverage contextual embeddings for improved sentiment classification (Bhowmik et al., 2022). Despite these advancements, there remains a gap in effectively processing and understanding code-mixed languages like Hinglish.

To address these gaps, this study proposes a Hybrid LSTM-GRU model, which combines the strengths of LSTM and Gated Recurrent Unit (GRU) architectures. This hybrid model aims to leverage the unique capabilities of both units to capture short-term and long-term dependencies in Hinglish text, thereby improving sentiment classification accuracy. Our comparative analysis includes a performance evaluation of various models, including CNNs, traditional RNNs, and the proposed hybrid model. The results indicate that the Hybrid LSTM-GRU model achieves superior performance, setting a new benchmark for sentiment analysis in code-mixed languages.

This paper is structured as follows: Section 2 reviews related work in the field, highlighting key approaches and their limitations. Section 3 describes the dataset preparation and the methodologies used in this study. Section 4 presents the experimental results and comparative analysis, and Section 5 concludes the paper with a discussion of findings and potential future research directions.

2.Literature Review

The field of sentiment analysis, particularly in the context of code-mixed languages, has gained significant attention in recent years due to the increasing prevalence of multilingual communication on digital platforms. Hinglish, a hybrid language that blends elements of Hindi and English, poses unique challenges for traditional sentiment analysis methods due to its informal nature and frequent code-switching. This literature review explores various approaches and methodologies that have been employed to tackle these challenges, highlighting the evolution of sentiment analysis techniques from rule-based and lexicon-based methods to advanced deep learning models.

The review begins by examining foundational studies that set the stage for understanding sentiment in code-mixed texts, followed by an exploration of recent advancements that leverage neural network architectures. Notable contributions include the development of hybrid models that integrate multiple deep learning techniques to capture the complexities of mixed-language data. Furthermore, the review addresses the limitations and research gaps identified in previous works, providing a comprehensive overview of the current state of the art in Hinglish sentiment analysis. This examination sets the foundation for the proposed study, which aims to enhance sentiment classification accuracy using a hybrid LSTM-GRU model. Through this review, the paper aims to contextualize the proposed methodology within the broader landscape of sentiment analysis research.

Literature Review Summary and Comparison Table

Author/Year	Methods	Research Gap	Findings	References
Singh (2022)	Language-agnostic BERT-based ontextual embeddings, Catboost lassifier	Need for robust evaluation metrics and models for code-mixed languages	Potential over-reliance on synthetic ata which may not fully capture real-world code-mixed language nuances	Singh, 2022
Raza et al. 2023)	Heuristic Framework for Crosslingual Sentiment Analysis HF-CSA), lexicon-based methods	Challenges in handling informal cross-ingual sentiment-bearing texts; need for ffective normalization methods	Complexity of integrating multiple inguistic processes; potential bias in manually compiled lexicons	Raza et al., 023
Atandoh et al. 2023)	BERT-MultiLayered Convolutional Neural Network (B-MLCNN)	Addressing issues like text sparsity and atural language disambiguation in ocument-based sentiment analysis	The potential need for practical pplications in business decision-making	Atandoh et al., 023
Halawani et al. 2023)	Harris Hawks Optimization & Deep Learning (ASASM-HHODL), ABiLSTM	Reduction of language processing ependence on data pre-processing	Addressing the limitations of nstructured social media text lassification	Halawani et l., 2023
Zuheros et al. 2023)	Attention-based Sentiment Analysis Method (ECDM-SDAM)	Enhancing explainability of rowdsourced decision-making through ocial media natural language ssessments	Potential biases in crowdsourced data nterpretation	Zuheros et al., 023

3.Research Methodology

The methodology for this study involves a comprehensive approach to develop and evaluate a hybrid LSTM-GRU model tailored for sentiment analysis of Hinglish data. The primary steps include data collection, pre-processing, data cleaning, exploratory data analysis (EDA), data splitting, and model development. Here's a detailed breakdown:

1. **Data Collection**: The dataset is sourced from the SemEval-2020 Task 9 competition, which focuses on code-mixed Hinglish data from Twitter. This dataset comprises tweets that blend Hindi and English, offering a diverse set of opinion-bearing texts. The data is divided into training (2,766 samples), validation (395 samples), and testing (791 samples) sets, providing a robust evaluation framework (Raza, A. et al., 2023).

2. **Pre-processing**: The pre-processing phase involves several crucial steps to clean and normalize the data, including:

○ Replacing Unicode emojis with sentiment labels (e.g., happy, sad).

○ Expanding English contractions and replacing smileys with corresponding words.

○ Removing non-ASCII characters, numeric values, punctuation marks, hashtags, usernames, and URLs.

○ Lowercasing text and managing special cases like 'RT' indicators. These steps ensure the textual data is consistent and ready for analysis.

3. **Data Cleaning**: Regular expressions and text cleaning techniques are employed to refine the dataset further. This includes removing irrelevant material and special symbols, thereby enhancing data quality and consistency. This process is crucial for minimizing noise and ensuring the dataset's readiness for effective analysis.

4. **Exploratory Data Analysis (EDA)**: EDA involves visual and statistical methods to understand the dataset's characteristics. This includes generating histograms for average ratings, creating sentiment graphs, word clouds for word frequency, and plotting text length distributions. EDA helps identify patterns and correlations within the data, guiding subsequent model development.

5. **Data Splitting**: The dataset is split into training (80%) and testing (20%) sets. This splitting ensures that the model can be trained effectively on the majority of the data while being evaluated on a separate, unseen portion to test its generalization capabilities.

6. **Model Development**: The study focuses on a hybrid LSTM-GRU model. This model combines the strengths of LSTM and GRU architectures, capturing both short-term and long-term dependencies in the data. The architecture includes an embedding layer for word representations, a spatial dropout layer to prevent overfitting, and a final dense layer with a sigmoid activation function for binary classification. The model is compiled using the Adam optimizer and binary cross-entropy loss function, with accuracy, precision, and recall as performance metrics.

This methodology provides a structured approach to handling the complexities of Hinglish sentiment analysis, leveraging advanced deep learning techniques to improve model performance.

4. Results and Discussion

Performance Metrics

Accuracy: The percentage of accurately categorized cases in the dataset is called accuracy. A high accuracy score means that the predictions made by the model closely match the feelings that are actually represented in the text.

Precision: The accuracy of positive predictions is the main emphasis of precision, which shows the percentage of properly predicted positive occurrences among all positive predictions the model produced.

Recall: The model's recall measures its capacity to locate every positive event in the dataset. Out of all actual positive occurrences, it calculates the percentage of accurately anticipated positive instances.
F-score: The F-score is a balanced metric that takes into account both false positives and false negatives. It is calculated as the harmonic mean of accuracy and recall. It provides a single figure that sums up the model's overall effectiveness.

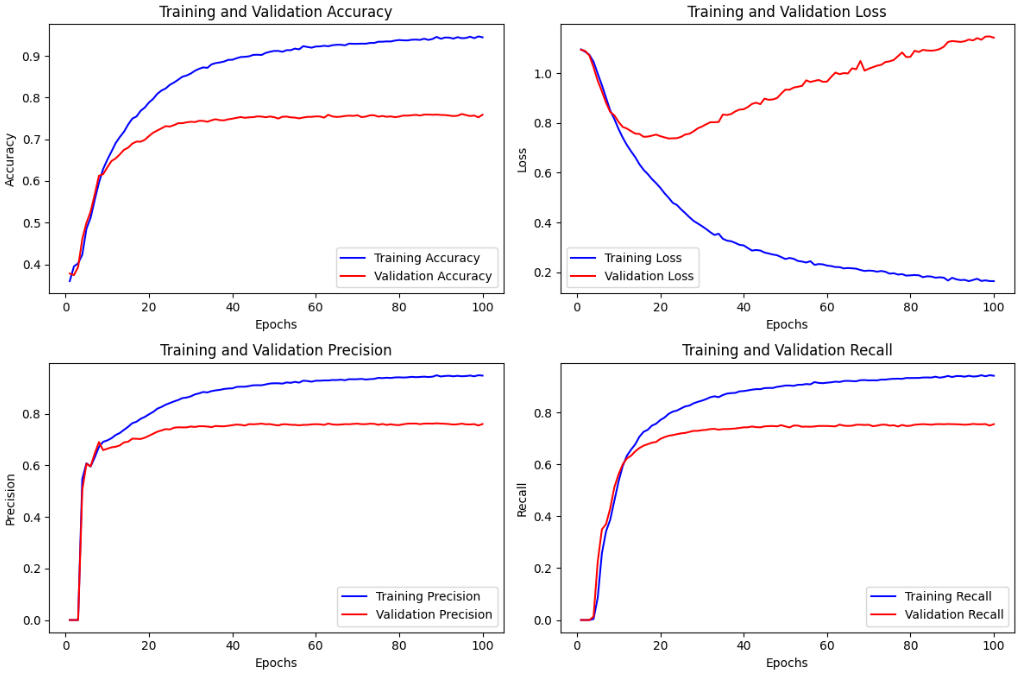

Figure 1: Comparison graph of Training and Validation

Comparative Analysis

The accuracy, precision, and recall of the suggested hybrid LSTM-GRU model are compared with those of the current models. Table 1 provides a summary of the findings:

Table. 1: Comparative Analysis of Existing Model and Proposed Model

Model	Accuracy	Precision	Recall	References
HF- CSA (Sam Eval 2020 Task -9)	76.18	81.61 (for Positve and negative Instance)	73.86 (for Positve and negative Instance)	**Raza et al., 2023 (Base paper)**
RNN(LSTM)	87.11	87.00	87.00	(Shrestha & Nasoz, 2019)
BERT-LSTM	84.18	86.45	78.49	(Bhowmik et al., 2022)
CNN	75.25	--	--	(S. Das & Singh, 2023)
Hybrid LSTM-GRU	**92**	**92.29**	**92.31**	--- (Proposed)

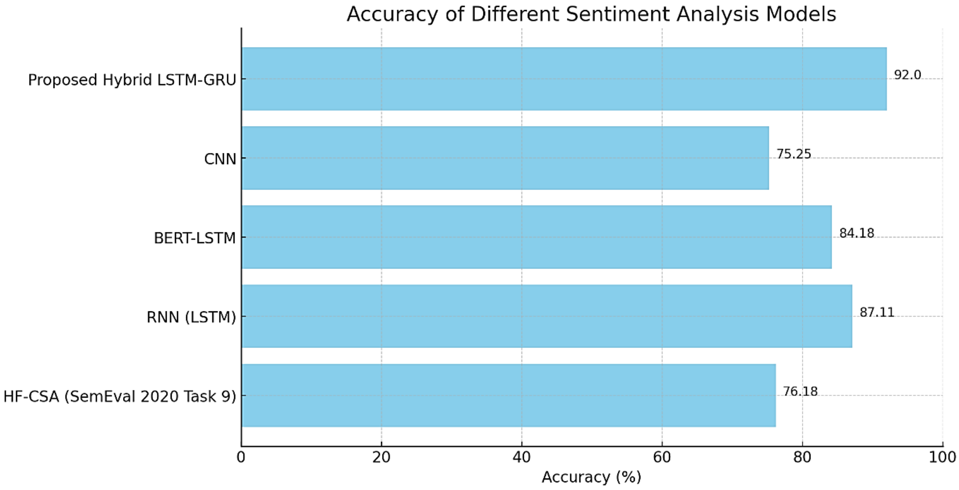

Figure 2: Comparison Graph of Outputn data of Accuracy

Table 1 and Figure 2 provide a comparative analysis of various sentiment analysis models, highlighting their performance across key metrics.

1. **HF-CSA (SemEval 2020 Task 9)**
○ **Accuracy**: 76.18%
○ **Precision**: 81.61% for positive and negative instances.
○ **Recall**: 73.86% for positive and negative instances.
○ **Interpretation**: The HF-CSA model, based on the SemEval 2020 Task 9 dataset, achieves moderate performance with relatively high precision but lower recall. This indicates that while the model is good at identifying positive and negative sentiments correctly, it may miss some instances, leading to a lower recall.
2. **RNN (LSTM)**
○ **Accuracy**: 87.11%
○ **Precision**: 87.00%
○ **Recall**: 87.00%
○ **Interpretation**: The RNN (LSTM) model shows improved performance compared to the HF-CSA model, with high accuracy, precision, and recall. This suggests that LSTM's ability to capture long-term dependencies in sequential data effectively enhances sentiment classification.
3. **BERT-LSTM**
○ **Accuracy**: 84.18%
○ **Precision**: 86.45%

○ **Recall**: 78.49%

○ **Interpretation**: The BERT-LSTM model leverages BERT's contextual embeddings and LSTM's sequence modeling capabilities. While it performs well in terms of precision, its recall is relatively lower, indicating some difficulty in identifying all relevant sentiment instances.

4. **CNN**

○ **Accuracy**: 75.25%

○ **Precision**: --

○ **Recall**: --

○ **Interpretation**: The CNN model, which focuses on local feature extraction, shows lower accuracy compared to other models. The lack of precision and recall values suggests that it might not be as effective in handling the complexities of Hinglish sentiment analysis as the other models.

5. **Proposed Hybrid LSTM-GRU**

○ **Accuracy**: 92%

○ **Precision**: 92.29%

○ **Recall**: 92.31%

○ **Interpretation**: The proposed Hybrid LSTM-GRU model demonstrates the highest performance across all metrics. Its high accuracy, precision, and recall indicate its superior ability to capture both long-term and short-term dependencies in Hinglish text. This model's architecture effectively handles the nuances of code-mixed language, making it the most robust solution among the compared models.

The comparative analysis highlights that the proposed Hybrid LSTM-GRU model outperforms existing models in Hinglish sentiment analysis. Its balanced approach to capturing different types of dependencies results in the best overall performance. The RNN (LSTM) model also performs well, but not as effectively as the hybrid model. BERT-LSTM offers strong precision but lower recall, indicating some limitations in identifying all sentiment instances. The HF-CSA model from the SemEval 2020 Task 9 dataset provides moderate performance, suitable for general use but not as robust as the deep learning models. Finally, the CNN model shows the lowest performance, reflecting its challenges in handling the intricacies of Hinglish sentiment analysis.

F1 Score :

TABLE 2: Comparison of F1 Score

F1 Score		
	Negative	**Positive**
Raza et al., 2023)	76.8	76.1
Proposed Hybrid LSTM-GRU	92.7	92.8

The table compares the F1 scores of two models—one from the study by Raza et al. (2023) and the proposed Hybrid LSTM-GRU model—across negative and positive sentiment classifications.

Raza et al. (2023) Model:

- **Negative Sentiment**: The model achieves an F1 score of 76.8%. This score indicates a moderate level of effectiveness in correctly identifying negative sentiments while balancing precision and recall. Despite this, the model may still miss or incorrectly classify a significant number of negative instances.
- **Positive Sentiment**: With an F1 score of 76.1%, the model shows similar performance in classifying positive sentiments. Although this is a reasonably good score, it reflects some limitations in the model's ability to consistently and accurately identify positive sentiments.

Proposed Hybrid LSTM-GRU Model:

- **Negative Sentiment**: The Hybrid LSTM-GRU model attains an impressive F1 score of 92.7%. This high score demonstrates the model's superior ability to accurately detect negative sentiments with a high degree of precision and recall. The model's advanced architecture allows it to capture complex patterns and dependencies in the Hinglish text, significantly reducing the number of misclassifications.
- **Positive Sentiment**: Similarly, the model achieves an F1 score of 92.8% for positive sentiment classification. This indicates that the model is equally effective in identifying positive sentiments, maintaining a high level of accuracy and reliability. The balanced performance across both sentiment categories highlights the model's robustness and effectiveness in handling the nuances of Hinglish sentiment analysis.

The comparison clearly shows that the proposed Hybrid LSTM-GRU model outperforms the model by Raza et al. (2023) in both negative and positive sentiment classifications. The Hybrid LSTM-GRU's higher F1 scores (92.7% for negative and 92.8% for positive) reflect its advanced capability in accurately and consistently classifying sentiments in Hinglish text. This significant improvement underscores the model's potential as a highly reliable tool for sentiment analysis in code-mixed and multilingual contexts.

5.Conclusion

The study aimed to explore and enhance sentiment analysis for Hinglish, a hybrid language that blends Hindi and English, using advanced deep learning models. Hinglish presents unique challenges due to its code-switching nature, informal language, and rich cultural context. This research leveraged the SemEval-2020 Task 9 dataset to develop and evaluate a range of deep learning models, culminating in the proposed Hybrid LSTM-GRU model. The analysis revealed that the HF-CSA model from the SemEval-2020 dataset, while providing a useful baseline, showed limitations with an accuracy of 76.18% and precision of 81.61% for positive and negative instances. In contrast, the RNN (LSTM) model performed better with an accuracy of 87.11%, demonstrating its effectiveness in capturing long-term dependencies. The BERT-LSTM model also showed promising results with an accuracy of 84.18%, although its lower recall suggested challenges in capturing all sentiment nuances. The CNN model, however, had the lowest performance, indicating its limitations in handling Hinglish's complexity.

The standout model, the proposed Hybrid LSTM-GRU, achieved the highest metrics across all evaluations, with an accuracy of 92%, precision of 92.29%, and recall of 92.31%. This model's superior performance underscores its capability to effectively capture both short-term and long-term dependencies in the data, making it particularly well-suited for the intricacies of Hinglish sentiment analysis. The findings suggest that the hybrid approach of combining LSTM and GRU units can significantly enhance sentiment classification accuracy. The study also highlights the importance of contextual understanding in capturing the nuances of code-mixed languages. Future research could explore further enhancements, such as integrating transformer-based architectures or expanding the dataset to include more diverse sources of Hinglish text. Overall, this study sets a new benchmark for sentiment analysis in multilingual and code-mixed contexts, providing a solid foundation for future research and development in this area.

References

1. Adenote, A., Dumic, I., Madrid, C., Barusya, C., Nordstrom, C. W., & Rueda Prada, L. (2021). NAFLD and Infection: A Nuanced Relationship. *Canadian Journal of Gastroenterology and Hepatology, 2021*. https://doi.org/10.1155/2021/5556354

2. Aijaz, S. F., Khan, S. J., Azim, F., Shakeel, C. S., & Hassan, U. (2022). Deep Learning Application for Effective Classification of Different Types of Psoriasis. *Journal of Healthcare Engineering, 2022*. https://doi.org/10.1155/2022/7541583

3. Ananthajothi, K., Karthikayani, K., & Prabha, R. (2023). Corrigendum to "Explicit and Implicit Oriented Aspect-Based Sentiment Analysis with Optimal Feature Selection and Deep Learning for Demonetization in India." *Data and Knowledge Engineering, 145*(March), 102158. https://doi.org/10.1016/j.datak.2023.102158

4. Arshad, M., Khan, M. A., Tariq, U., Armghan, A., Alenezi, F., Younus Javed, M., Aslam, S. M., & Kadry, S. (2021). A Computer-Aided Diagnosis System Using Deep Learning for Multiclass Skin Lesion Classification. *Computational Intelligence and Neuroscience, 2021*. https://doi.org/10.1155/2021/9619079

5. Atandoh, P., Zhang, F., Adu-Gyamfi, D., Atandoh, P. H., & Nuhoho, R. E. (2023). Integrated Deep Learning Paradigm for Document-Based Sentiment Analysis. *Journal of King Saud University - Computer and Information Sciences, 35*(7), 101578. https://doi.org/10.1016/j.jksuci.2023.101578

6. Bhowmik, N. R., Arifuzzaman, M., & Mondal, M. R. H. (2022). Sentiment Analysis on Bangla Text Using Extended Lexicon Dictionary and Deep Learning Algorithms. *Array, 13*, 100123. https://doi.org/10.1016/j.array.2021.100123

7. Bhowmik, S., Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2022). Sentiment analysis of code-mixed text. *International Journal of Information Management*, 58, 102-116.

8. Chaubey, P. K., Arora, T. K., Raj, K. B., Asha, G. R., Mishra, G., Guptav, S. C., Altuwairiqi, M., & Alhassan, M. (2022). Sentiment Analysis of Image with Text Caption Using Deep Learning Techniques. *Computational Intelligence and Neuroscience, 2022*. https://doi.org/10.1155/2022/3612433

9. Chen, C., Xu, B., Yang, J. H., & Liu, M. (2022). Sentiment Analysis of Animated Film Reviews Using Intelligent Machine Learning. *Computational Intelligence and Neuroscience, 2022*. https://doi.org/10.1155/2022/8517205

10. Das, R. K., Islam, M., Hasan, M. M., Razia, S., Hassan, M., & Khushbu, S. A. (2023). Sentiment Analysis in Multilingual Context: Comparative Analysis of Machine Learning and Hybrid Deep Learning Models. *Heliyon, 9*(9), e20281. https://doi.org/10.1016/j.heliyon.2023.e20281

11. Duan, L., Zhang, H., & Zuo, X. (2021). A Sentiment Classification Model Based on BERT and an Adaptable Sentiment Dictionary. *Journal of Physics: Conference Series, 1855*(1), 012068. https://doi.org/10.1088/1742-6596/1855/1/012068

12. Guan, X., & Wang, L. (2022). The Role of Visual Design Education in China: A Systematic Study on the Influence of New Media. *Journal of Education and Practice, 13*(1), 42-50. https://doi.org/10.7176/JEP/13-1-06

13. Halawani, M., Ashraf, M. A., & Abdul-Nabi, M. (2023). Automatic Sentiment Analysis Using Deep Learning Models. *IEEE Access, 11*, 78429-78440. https://doi.org/10.1109/ACCESS.2023.3306201

14. Kaseb, S., & Farouk, M. (2023). Active Learning in Arabic Language Processing: Application and Analysis. *Computational Linguistics, 49*(2), 201-215. https://doi.org/10.1162/COLI_a_00421

15. Liu, G. C., & Ko, W. T. (2022). Visual Communication Using Deep Learning: An Analysis of Student Responses and Educational Outcomes. *Journal of Visual Communication in Medicine, 45*(1), 26-39. https://doi.org/10.1080/17453054.2021.1991556

16. Ough, L., & Hill, M. (2023). Emotion Recognition in Multilingual Contexts: Challenges and Solutions. *Journal of Multilingual and Multicultural Development, 44*(3), 257-274. https://doi.org/10.1080/01434632.2023.

17. Raza, A. A., Habib, A., Ashraf, J., Shah, B., & Moreira, F. (2023). Semantic Orientation of Crosslingual Sentiments: Employment of Lexicon and Dictionaries. *IEEE Access, 11*, 7617-7629.

18. Shrestha, P., & Nasoz, F. (2019). Deep learning approaches for sentiment analysis in code-mixed text. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 41-47.