VAWKUM Transactions on Computer Sciences
http://vfast.org/journals/index.php/VTCS@ 2025, ISSN(e): 2308-8168, ISSN(p): 2411-6335
Volume 13, Number 1, January-June 2025 pp: 190-200

Sentiment Analysis of Balochi Text Using Deep Learning

Shumaila Hussain [1], **Sibghat Ullah Bazai** [2 *], **Shahab Qadir** [3], **Shah Marjan** [4], **Muhammad Imran Ghafoor** [5], **Paras Pervaiz** [3]

[1]Department of Computer Science Sardar Bahadur Khan Women's University, Quetta Pakistan; [2]Department of Computer Engineering Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan; [3]Department of Computer Science, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan ; [4]Department of Software Engineering Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan; [5]Department of Engineering, Pakistan Television Corporation, Lahore, Pakistan

Keywords: *Deep Learning, Balochi Text Classification , Balochi Dataset, Sentiment Analysis*

Journal Info:
Submitted:
February 26, 2025
Accepted:
March 30, 2025
Published:
May 09, 2025

Abstract

Balochi is a low-resource language with limited available data for computational modelling. This study aims to perform sentiment analysis on Balochi text using machine learning techniques. To address the scarcity of linguistic data, we contribute a large, newly constructed dataset of Balochi text. Our proposed model incorporates feature extraction and data augmentation within deep learning algorithms to classify sentiments as positive, negative, or neutral. We evaluate both traditional machine learning methods—such as Random Forest and Support Vector Machine (SVM)—and advanced deep learning models, including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU).The experimental evidence proves that LSTM and GRU are more effective than traditional methods, and the accuracy rates of sentiment classification with their help are 83.57% and 81.23%, respectively. It has been experimentally verified that, when it comes to the Balochi sentiment analysis, deep learning methods can be more effective than the traditional ones.

***Correspondence author email address:** sibghat.ullah@buitms.edu.pk
DOI: 10.21015/vtcs.v13i1.2081

1 Introduction

Sentiment analysis is very important in gauging the voiced opinions and the feelings of people about products, services, brands or socio-political concerns. As a subfield within the wider text categorization process, sentiment analysis is concerned with the process of determining and decoding the emotion or feeling that a text provides.

Social media sentiment analysis is an emerging trend in measuring the opinion of the people in different languages. Sentiment analysis through the application of machine learning methods has been demonstrated to be very successful in finding insights, as well as meaningful information, in textual data because of the strong text classification abilities of machine learning. Although some algorithms of traditional text classification do tend to look at the objective characteristics of a text, like the topic or the author sentiment analysis method does specifically target the subjective ones, such as, e.g., the tone of emotion, context, and the general state of mind that the text conveys. Due to this sentiment analysis necessitates the application of special features and methods, which are not similar to those employed in traditional text categorization [1].

Sentiment analysis can be done using a wide range of machine learning models and algorithms, among them rule-based, Naive Bayes, support vector machine (SVM), and deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) and specifically Long Short-Term Memory (LSTM) networks. Such algorithms are normally trained using annotated text sets, and every sentence is categorized in terms of its sentiment or polarity rating. Their performance is measured through usual metrics like precision, accuracy, F1-score, and recall as [2].

Sentiment analysis is not properly developed in most of the widely spoken languages despite the fact that it is a prominent field of research. A good example is the Balochi language. Balochi is the main language of the Baloch people inhabiting the Balochistan region, and communities all over the Middle East, Europe and the rest of the world, with an estimated 10 million speakers of the language. Although the Balochi language adheres to the scheme of the Arabic script, it has its own system of writing comprising 29 letters[3] associated with it as shown in Figure 1.

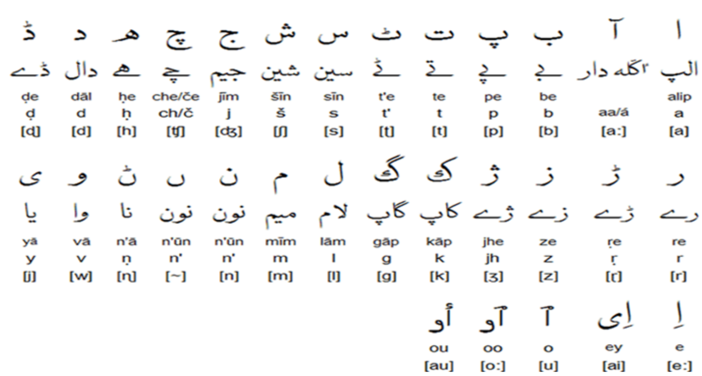

Figure 1. Balochi Alphabets

Balochi has been regarded as one of the low resource languages to be used in machine learning. Its non-formal and heterogeneous textual characteristics have traditionally introduced considerable difficulties to the natural language processing (NLP) tasks. The given study is supposed to make a contribution as it will examine sentiment analysis of the Balochi text, and, as far as we know, this aspect has not been studied yet. Findings of the study will allow native Balochi speakers to become more active in the online environment and enjoy the improvement of language technology.

2 Literature Review

Deep learning is a particular type of artificial intelligence (AI), which emulates the neural structure of the human brain in its ability to handle complex information. Deep learning methods have been widely used in a variety of fields, such as image recognition, plant disease detection, and text classification [4, 5]. Convolutional Neural Networks (CNN), Long-Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) are deep learning models, which

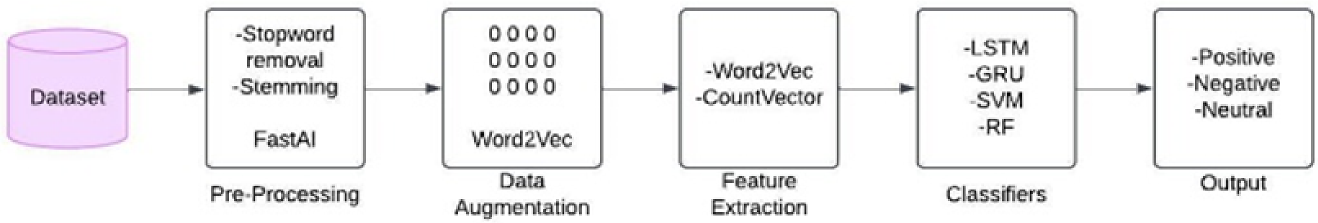

Figure 2. Research Methodology

have shown impressive success in text classification tasks in natural language processing (NLP). As an illustration, these algorithms have been benchmarked to process Hindi language to determine the most efficient architecture to achieve the best performance.

In addition to deep learning, other techniques of machine learning, such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, Decision Trees, and Random Forest (RF) have also been employed in the study of text categorization studies [6, 7]. Equally, in another research, various deep learning models were tested and the transfer learning method of Bi-LSTM was tested to be superior to sentiment analysis of low-resource Persian language [8]. Arabic text Sentiment analysis A second study researched sentiment analysis of arabic text and found no method that performed better than SCM [9].

The Support Vector Machine (SVM) model proved to be more efficient and reached the highest accuracy of 54.932% in case of a Sindhi text corpus. This paper highlights the importance of the quality text corpora in the supplementation of different natural language processing (NLP) activities, such as automated translation, sentiment analysis, text classification, and information extraction [10].

A Convolutional Neural Network (CNN) was used to examine sentiment in Arabic social media text through Arabic Sentiment Tweets Dataset (ASTD) in a parallel study. The model also included feature extraction using machine learning algorithms, which finally recorded an accuracy of 64.30% and 64.09% F1-score proving the use of CNN in the Arabic sentiment analysis process is effective [11–13]. The current literature has used deep learning methods to analyze sentiment in other areas. In the case of Persian-language e-commerce reviews, the researchers used Convolutional Neural Networks (CNN) and Long Short-Memory (LSTM) networks [14]. Correspondingly, comparative studies of CNN, LSTM, Transformer encoders, BERT, and Gated Recurrent Units(GRU) showed that CNN performs better in medical record analysis, although these two models are closely related to each other in other types of research tasks or tasks when it comes to medical record analysis [15].

Sentiment analysis has been widely used on South Asian languages like Urdu, Pashto, and Sindhi with the use of deep learning algorithms like SVM, LSTM, and CNN [16–25]. Notably, research on Roman Urdu text from social media platforms revealed that the Word2Vec Continuous Bag-of-Words (CBOW) framework combined with SVM outperformed both BERT word-embedding SVM and two-layer LSTM architectures in emotion classification [26–31]. Similarly, the other application areas including product pricing strategies, customer retention analysis, brand development, and targeted advertising LSTM, CNN, and GRU proved efficiency [32][33].

We have focused on Balochi sentiment analysis and faced challenges in dataset creation, character recognition etc. the proposed model implemented feature extraction and data augmentation to address dataset imbalance significantly improved the performance of machine learning models for sentiment analysis. This research represents an important step in preserving Balochi language presence in digital spaces through computational analysis of native vocabulary sentiment.

3 Materials and Methods

The research flowchart is depicted in Figure 2, illustrating the steps involved in this study. The first step involves dataset collection, which was scrupulously gathered from internet resources, including short stories, online articles, and websites. This was necessary because there aren't many internet resources available for the Balochi language. The second step is pre-processing the dataset, which includes stop word removal and stemming using FastAI/Spacy. In the third step, data augmentation is performed using Word2Vec to balance the dataset.

In the fourth step, feature extraction is carried out. The feature retrieval and representation approaches used to convert textual data into numerical form suitable for the models are explained. For feature extraction, both CountVectorizer and Word2Vec were utilized. While CountVectorizer converts the text into a matrix of token counts, Word2Vec employs a neural network to map words to vectors for Balochi language text classification. In the fifth step, classification is performed. The data is categorized into groups for evaluation, verification, and training to assist in the creation and assessment of the framework. The Sentiment analysis classification is carried out using three categories: neutral, negative, and positive attitudes. The output represents the text into selected sentiments.

For Balochi Language text classification, the Long Short Term Memory, Random Forest, Gated Recurrent Units, and Support Vector Machine (SVM) models were chosen. These models were trained with hyper-parameters, data splitting, and augmentation to balance the class distribution of the training data. The models were assessed using performance criteria like accuracy, precision, recall, and F1-score.

3.1 Dataset

As a result of the limited resources available on the Balochi language, we have created a dataset containing 1,701 phrases, tokenized to 386,516 words that are collected in a variety of sources. It has been categorized into three sentiments to analyse them, namely, neutral, negative, and positive.

3.2 Dataset Pre-Processing

The dataset follows an extensive pre-processing to make it more appropriate to machine learning algorithms. Using the FastAI and the tokenizer of Spacy, we have done stemming to make the text more compact.

3.3 Data Augmentation

The dataset is augmented with the use of Word2Vec technique to resolve the problem of data imbalance.

3.4 Data Segmentation

Using the train test split, the dataset is divided into the training and test sets to have an accurate model evaluation. The training set (X train and y train) captures eighty percent and the remaining twenty percent is captured by the testing set (X test and y test). To have repeatability, this split was done arbitrarily but a fixed seed (random state=42).

3.5 Feature Extraction

The extraction of features is done in such a way that the maximum amount of information that is important is included hence improving the execution of machine learning algorithms. In the present research, Word2Vec and CountVectorizer methods have been used in the extraction of features. Word2Vec is a neural network model that converts text into the form of vectors that reflect the semantic meaning of each word. An example of the feature extraction procedure is illustrated in Figure 3. Countvectorizer, on the other hand, produces a list of counts of tokens depending on the given text. A vocabulary package, which captures the content of the dataset, was created as a result of this inquiry using CountVectorizer.

Machine learning algorithms are used to extract numerical features of the dataset which are used as inputs to the model to perform sentiment analysis on Balochi text. These are the numbers that make the algorithms

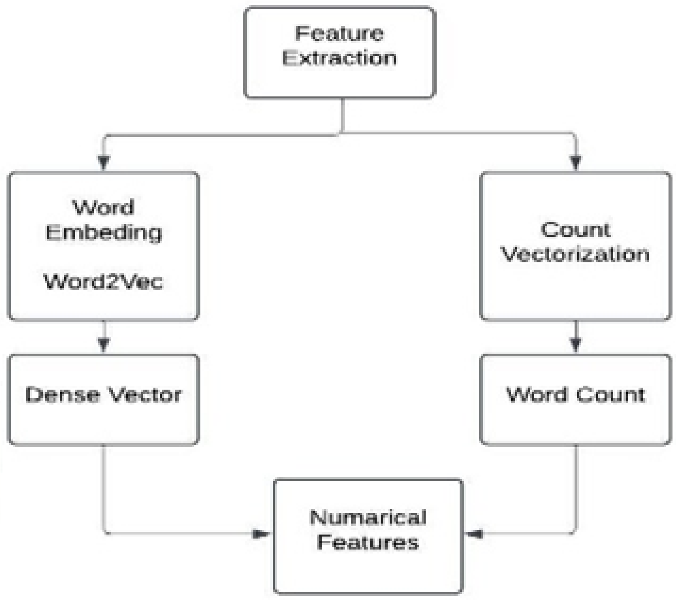

Figure 3. Feature Extraction

detect and regulate the underlying sentiments and lead to effective classification and appraisal of content that is emotionally instigated or opinion-based.

3.6 Sentiment Analysis

Machine learning algorithms receive input in the processed Balochi text and exhibit such attributes. These algorithms assign sentiment tags, which put the text in three categories of neutral, negative and positive after analyzing the narrative data. The categorizing process ensures that the sentiment predictions are accurate through the structured nature of the data that has been cleaned. In this study, the sentiments in Balochi text have been studied and analyzed through four text categorization techniques. Deep learning models such as Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) are likely especially effective in text categorization. Random Forest, Support Vector Machine (SVM), and Long Short Term Memory (LSTM) are utilized because they can detect dependencies that can last. This work employed methods of optimization and hyperparameter values turnings to train the models. This was carried out in order to guarantee that the models were trained as efficiently as possible. Using a batch capacity of 32, the models were programmed for a maximum of 50 epochs for the Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM). A comprehensive overview of the sentiment dispersion in the Balochi text dataset is provided by the following Table 1, which illustrates the sentiment analysis results based on by the machine learning algorithms. The text was evaluated by separating evaluations into neutral, negative and positive.

4 Result and Discussion

4.1 Evaluation Metrics

The results were evaluated using the most impactful measures while dealing with deep learning algorithms including precision, recall, F1 and accuracy.

4.1.1 Precision

It is used to describe the proportion of values that actually have a problem. Mathematically, it looks like this:

Table 1. Sentiment Analysis of Balochi Text

Positive Reviews	Negative Reviews	Neutral Reviews
بانک کارینا جھانی بلوچی ء سے مزنیں گالورانی تہ ء بھر ء بانگ کنگ ء ھم نکشون ء وسیلہ ء پیش دارایت	اے ساھت ء امینچو گپ بس انت کہ کاگد مزن بوت بگندئے اے جست بے سریں جست بہ بنت	ہجہ بوت بگندئے اے جست بے سریں جست بہ بنت
من ءَ اے نبشتانک سک دوست بیت	تو انچو گپ ء ئے گش ئے ننک ایوک ء تئیگ انت ء من ئی ء دزگ ء اتکگاں من بچکندت	من اے کتاب شیلف سرا دیستگ

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

4.1.2 Recall

The measurement of the proportion of real cases actually predicted by the model. Mathematically, it is represented as below:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

4.1.3 F1 Score

The F1 score is calculated as the harmonic mean of the precision and recall scores. Mathematically it is represented as

$$F1 = \frac{TP}{Tp + \frac{1}{2(FP+FN)}}$$

(3)

4.2 Long-Short-Term-Memory

The performance of Long Short-Term Memory model is depicted in Table 2. it displays its precision which is 87.58% while recall of 79.66%, the F1 score of 83.43% while the accuracy of 83.58% . It signifies the overall prediction power of LSTM for sentiment analysis of Balochi text.

Table 2. LSTM Classification Report for Balochi Sentiment Analysis

Classifier Metrics	Results
Precision	87.57%
Recall	79.66%
F1-Score	83.43%
Accuracy	83.57%

4.3 Gated Recurrent Units (GRU)

The precision score of GRU is 86.93%, recall is 75.14%, F1 score is 80.61% while accuracy is 81.23%. The Table 3 indicated the performance of GRU model.

Table 3. GRU Classification Report for Balochi Sentiment Analysis

Classifier Metrics	Results
Precision	86.92%
Recall	75.14%
F1-Score	80.60%
Accuracy	81.23 %

4.4 Support Vector Machine (SVM)

The SVM model exhibits a commendable accuracy rate of 88% for the Balochi Text. The precision score of 88.28%, recall of 87.93%, F1 score of 87.67%. The overall performance of SVM is depicted in Table 4. The SVM was better because the imbalance in the data has been corrected by the study through data augmentation that made it more appropriate and enabled to enhance its capacity of handling high dimensional data, to pick on non linear relationship as well as resistance to noise.

Table 4. Support Vector Machine Classification Report for Balochi Sentiment Analysis

Classifier Metrics	Results
Precision	88.28%
Recall	87.93%
F1-Score	87.67%
Accuracy	88.0 %

4.5 Random Forest

The accuracy of Random Forest is estimated at 82.70%, its precision score is 98.36%, its recall is 67.80 per cent and F1-score is 80.27%. This generic performance of random forest is shown in Table 7. To categorize the Balochi text as either positive, negative or neutral remarks.

Table 5. Random Forest Report for Balochi Sentiment Analysis

Classifier Metrics	Results
Precision	93.36%
Recall	67.79%
F1-Score	80.26%
Accuracy	82.69 %

4.6 Bidirectional Long Short-Term Memory

The BILSTM was showing 77.70% accuracy, precision score of 75 percent, recall of 77.80% and F1 score is 76.27%. According to the Table 7, the performance of the entire model is shown to determine the Balochi sentiments as positive, negative or neutral.

Table 6. Bidirectional Long Short-Term Memory for Balochi Sentiment Analysis

Classifier Metrics	Results
Precision	75.06%
Recall	77.0%
F1-Score	76.6%
Accuracy	77.08 %

4.7 Semantic Clustering Mechanism

The Semantic Clustering Mechanism SCM demonstrated 85.00%, accuracy with 85 % precision score of 89%, recall of 88.0%, and F1 score is 87.00%. The Table 7 indicates the overall performance of the model to classify the sentiments of Balochi as positive, negative or neutral.

Table 7. Semantic Clustering Mechanism for Balochi Sentiment Analysis

Classifier Metrics	Results
Precision	89.00%
Recall	88.0%
F1-Score	87.0%
Accuracy	88.08 %

The following Figure 4 illustrates the general efficiency of some of the specified deep learning algorithms.

The results demonstrate the effectiveness of deep learning models and establish benchmarks for sentiment analysis of Balochi text. The selected machine learning algorithms included Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), Support Vector Machines (SVM), and Random Forest. Among these, SVM outperformed the others when evaluated using metrics such as accuracy, F1 score, precision, and recall. The SVM had an accuracy of 88 per cent and an F1 score of 87%. Although the LSTM demonstrated a relatively low accuracy and F1 score of 83% as in Table 6. Based on the outcomes, SVM is the leading algorithm to use as a sentiment analysis of Balochi text.

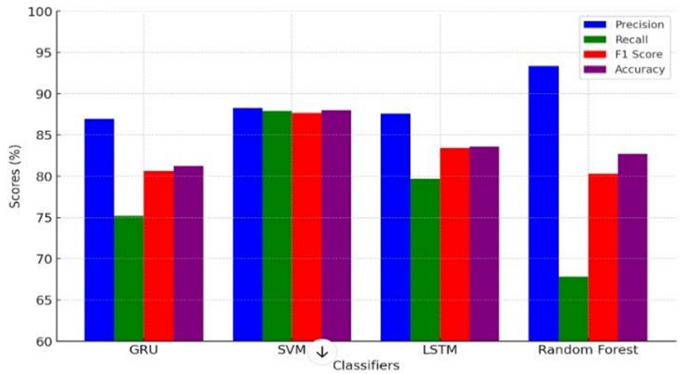

Figure 4. Performance Comparison of Deep Learning Algorithms figure

5 Conclusion and Future Work

The given research considers the severe problem of conducting sentiment analysis and text classification in Balochi, a language of great linguistic value, but computationally under-resourced. In order to fill this gap, we add a multi-source, multi-pre-processed Balochi text dataset to be used in machine learning. The curated corpus is useful in the future research and practical implementations in Balochi language processing.

To improve sentiment analysis of low-resource Balochi text, our methodology uses the new feature extraction methods, and data augmentation to overcome the class imbalance. We perform a systematic review of what is presented as the state-of-the-art deep learning architectures (Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU)) upon comparisons with traditional machine learning models (Support Vector Machines (SVM) and Random Forest). In this comparative analysis, the technique of optimizing sentiment classification has been proposed using our technique.The work has important implications in the new sphere of Balochi computational linguistics as it addresses the main problem of sentiment classification, and offers the evidence of the relative efficiency of modern and traditional analytical methods with our suggested technique. Both the findings of the research help to set certain valuable standards in Balochi language processing and to prove the flexibility of rich NLP methods to low-resource conditions.

Author Contributions

Shumaila Hussain: Writing- Reviewing and Editing, Investigation. **Sibghat Ullah Bazai**: Conceptualization, Methodology, Software **Shahab Qadir**: Data curation, Writing- Original draft preparation. **Muhammad Imran Ghafoor**: Visualization **Shah Marjan**: Supervision.: **Paras Pervaiz**: Software, Validation.

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

References

[1] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. J. Passonneau, "Sentiment analysis of twitter data," in Proceedings of the workshop on language in social media (LSM 2011), pp. 30–38, 2011.

[2] X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting bert for end-to-end aspect-based sentiment analysis," arXiv preprint arXiv:1910.00883, 2019.

[3] C. Jahani, A grammar of modern standard Balochi. Acta Universitatis Upsaliensis, 2019.

[4] K. Melcher, “A friendly introduction to [deep] neural networks,” 2021.

[5] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep learning–based text classification: a comprehensive review,” ACM computing surveys (CSUR), vol. 54, no. 3, pp. 1–40, 2021.

[6] R. Joshi, P. Goel, and R. Joshi, “Deep learning for hindi text classification: A comparison,” in Intelligent Human Computer Interaction: 11th International Conference, IHCI 2019, Allahabad, India, December 12–14, 2019, Proceedings 11, pp. 94–101, Springer, 2020.

[7] X. Hu and R. Zhang “Text Classification based on Machine Learning” in 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), pp. 911–916, IEEE, 2022.

[8] J. F. Ruma, S. Akter, J. Laboni, and R. M. Rahman " A deep learning classification model for Persian Hafez poetry based on the poet’s era" in Decision Analytics Journal, vol. 4, Science Direct,2022.

[9] A. Hawalah " Semantic Ontology-Based Approach to Enhance Arabic Text Classification" in big data and cognitive computing, MDPI, 2019.

[10] M. A. Dootio and A. I. Wagan, “Development of sindhi text corpus,” Journal of King Saud University-Computer and Information Sciences, vol. 33, no. 4, pp. 468–475, 2021.

[11] M. Heikal, M. Torki, and N. El-Makky, “Sentiment analysis of arabic tweets using deep learning,” Procedia Computer Science, vol. 142, pp. 114–122, 2018.

[12] A. M. Alayba, V. Palade, M. England, and R. Iqbal, “A combined cnn and lstm model for arabic sentiment analysis,” in Machine Learning and Knowledge Extraction: Sec- ond IFIP TC 5, TC 8/WG 8.4, 8.9, TC 12/WG 12.9 International Cross-Domain Con- ference, CD-MAKE 2018, Hamburg, Germany, August 27–30, 2018, Proceedings 2, pp. 179–191, Springer, 2018.

[13] A. Elnahas, N. Elfishawy, M. Nour, and M. Tolba, “Machine learning and feature selection approaches for categorizing arabic text: Analysis, comparison, and proposal,” The Egyptian Journal of Language Engineering, vol. 7, no. 2, pp. 1–19, 2020.

[14] Z. B. Nezhad and M. A. Deihimi, “A combined deep learning model for persian sentiment analysis,” IIUM Engineering Journal, vol. 20, no. 1, pp. 129–139, 2019.

[15] H. Lu, L. Ehwerhemuepha, and C. Rakovski, “A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance,” BMC Medical Research Methodology, vol. 22, no. 1, p. 181, 2022.

[16] I. U. Khan, A. Khan, W. Khan, M. M. Su’ud, M. M. Alam, F. Subhan, and M. Z. Asghar, “A review of urdu sentiment analysis with multilingual perspective: A case of urdu and roman urdu language,” Computers, vol. 11, no. 1, p. 3, 2022.

[17] W. Ali, N. Ali, Y. Dai, J. Kumar, S. Tumrani, and Z. Xu, “Creating and evaluating re- sources for sentiment analysis in the low-resource language: Sindhi,” in Proceedings of the Eleventh Workshop on Computational Approaches To Subjectivity, Sentiment and Social Media Analysis, pp. 188–194, 2021.

[18] H. Elfaik and E. H. Nfaoui, “e,” Journal of Intelligent Systems, vol. 30, no. 1, pp. 395– 412, 2020.

[19] S. Iqbal, F. Khan, H. U. Khan, T. Iqbal, and J. H. Shah, “Sentiment analysis of social media content in pashto language using deep learning algorithms,” Journal of Internet Technology, vol. 23, no. 7, pp. 1669–1677, 2022.

[20] F. Barakzai, S. Bhatti, and S. Saddar, “Sentiment analysis of sindhi news articles using deep learning,” in 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), pp. 26–31, IEEE, 2022.

[21] K. I. Malik, “Urdu news content classification using machine learning algorithms,” Lahore Garrison University Research Journal of Computer Science and Information Technology, vol. 6, no. 1, pp. 22–31, 2022.

[22] U. Naqvi, A. Majid, and S. A. Abbas, “Utsa: Urdu text sentiment analysis using deep learning methods,” IEEE Access, vol. 9, pp. 114085–114094, 2021.

[23] L. Khan, A. Amjad, N. Ashraf, H.-T. Chang, and A. Gelbukh, “Urdu sentiment anal- ysis with deep learning methods,” IEEE Access, vol. 9, pp. 97803–97812, 2021.

[24] X. Luo, “Efficient english text classification using selected machine learning tech- niques,” Alexandria Engineering Journal, vol. 60, no. 3, pp. 3401–3409, 2021.

[25] A. Setyanto, A. Laksito, F. Alarfaj, M. Alreshoodi, I. Oyong, M. Hayaty, A. Alomair,N. Almusallam, and L. Kurniasari, “Arabic language opinion mining based on long short-term memory (lstm),” Applied Sciences, vol. 12, no. 9, p. 4140, 2022.

[26] L. Khan, A. Amjad, K. M. Afaq, and H.-T. Chang, “Deep sentiment analysis using cnn-lstm architecture of english and roman urdu text shared in social media,” Applied Sciences, vol. 12, no. 5, p. 2694, 2022.

[27] A. C. M. V. Srinivas, C. Satyanarayana, C. Divakar, and K. P. Sirisha, “Sentiment analysis using neural network and lstm,” in IOP conference series: materials science and engineering, vol. 1074, p. 012007, IOP Publishing, 2021.

[28] F. Aftab, S. U. Bazai, S. Marjan, L. Baloch, S. Aslam, A. Amphawan, and T.-K. Neo, “A comprehensive survey on sentiment analysis techniques,” *International Journal of Technology*, vol. 14, no. 6, pp. 1288–1298, 2023.

[29] S. Fakhar, J. Baber, S. U. Bazai, S. Marjan, M. Jasinski, E. Jasinska, M. U. Chaudhry, Z. Leonowicz, and S. Hussain, “Smart classroom monitoring using novel real-time facial expression recognition system,” *Applied Sciences*, vol. 12, no. 23, p. 12134, 2022, doi: 10.3390/app122312134.

[30] A. Topbaş, A. Jamil, A. A. Hameed, S. M. Ali, S. Bazai, and S. A. Shah, “Sentiment analysis for COVID-19 tweets using recurrent neural network (RNN) and bidirectional encoder representations (BERT) models,” in *2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, Quetta, Pakistan, 2021, pp. 1–6, doi: 10.1109/ICECube53880.2021.9628315.

[31] I. Nasir, S. U. Bazai, M. I. Ghafoor, and S. Marjan, “Urdu sentiment analysis using deep learning,” in *2024 18th International Conference on Open Source Systems and Technologies (ICOSST)*, Dec. 2024, pp. 1–7.

[32] J. Singh and P. Tripathi, “Sentiment analysis of Twitter data by making use of SVM, random forest and decision tree algorithm,” in *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, 2021, pp. 193–198.

[33] Z. Zaland, S. U. Bazai, M. I. Ghafoor, S. Akram, M. Waseem, and M. Qaseem, “Vote bank forecast based on social media sentiment analysis,” in *2023 17th International Conference on Open Source Systems and Technologies (ICOSST)*, Dec. 2023, pp. 1–8.