Preprints.org

Review

Multilingual Sentiment Analysis Using Deep Learning: Survey

Uswa Ihsan , Humaira Ashraf * , NZ Jhanjhi

Review

Multilingual Sentiment Analysis Using Deep Learning: Survey

Uswa Ihsan [1], Humaira Ashraf [1] and NZ Jhanjhi [2]

[1] Department of Computer Science, Faculty of Computing; Information Technology, International Islamic University Islamabad

[2] School of Computer Science, SCS, Taylors University Malaysia, uswaihsan.mscs1075@iiu.edu.pk, humaira.ashraf@iiu.edu.pk,   noorzaman.jhanjhi@taylors.edu.my

Abstract: The rise of the Internet has enabled people to express their opinions on various subjects through social media, blogs, and website comments. As a result, there has been a significant increase in research on sentiment analysis. However, most of the research efforts have focused on analyzing sentiment in English-language data, neglecting the wealth of information available in other languages. In this paper, we provide a comprehensive review of the current state-of-the-art in multilingual sentiment analysis. The survey investigates techniques for data preprocessing, representation learning, and feature extraction in multilingual sentiment analysis. It explores cross-lingual transfer learning, domain adaptation, and data augmentation methods that enhance the performance of sentiment analysis models, particularly in low-resource languages and domains. It provides insights into the state-of-the-art approaches, challenges, and opportunities in this evolving field and encourages further advancements in multilingual sentiment analysis research.

Keywords: sentiment; deep learning; multilingual

Introduction

In order to construct a multilingual sentiment analysis, this study will examine various combinations of preprocessing methods, approaches of sentiment analysis, and assessment models that have been used in previously presented models. Additionally, cross-lingual SER is crucial in real-world applications, particularly when users from various cultural and linguistic backgrounds communicate with the system. The plethora of research on English SA, lexical resources, and natural language processing (NLP) techniques has significantly improved and successful. In comparison, substantially more work is required to reach the same level of performance in other languages, such as Arabic, Urdu, and French [1].

Different Surveys and reviews are conducted on multilingual sentiment analysis. The survey has covered the state of the art of sentiment analysis research for the Portuguese language describe and classify works that broadly relate to sentiment analysis [2]. Discussing works using sentiment analysis techniques tailored to the Portuguese language and its auxiliary materials, such as natural language processing (NLP) tools, lexicons corpora, ontologies, taxonomies, and databases. There is a problem   with the evaluation model chosen for the analytical process, according to the literature [3]. There is no particular method for multilingual sentiment analysis evaluation. It indicates that the evaluation model was chosen without consideration of academic merit and that the researcher alone chose the evaluation criteria. Despite the widespread use of these evaluation models in practice, it is crucial to identify a set of general assessment standards that can accept different languages without introducing biases towards particular datasets.

We have presented a systematic Literature review by reviewing different combination of studies for multilingual sentiment Analysis since 2019-2023 by using IEEE, Google scholar, ACM Databases. A search protocol was created, and thorough searches were carried out in accordance with it. The creation of these search strings was guided by the specific research questions of the study. Subsequently, a search strategy was employed to categorize all the conducted searches. Following the journal search, the research papers underwent a filtering process based on inclusion criteria, as

well as an assessment of their title, abstract, and objectives. The selected research papers were then further filtered based on their relevance to the research query as determined by their title and abstract. We have discussed objective and techniques for each paper thoroughly. At last we have performed critical analysis for each research study. Figure 1 shows the organization of our research review.
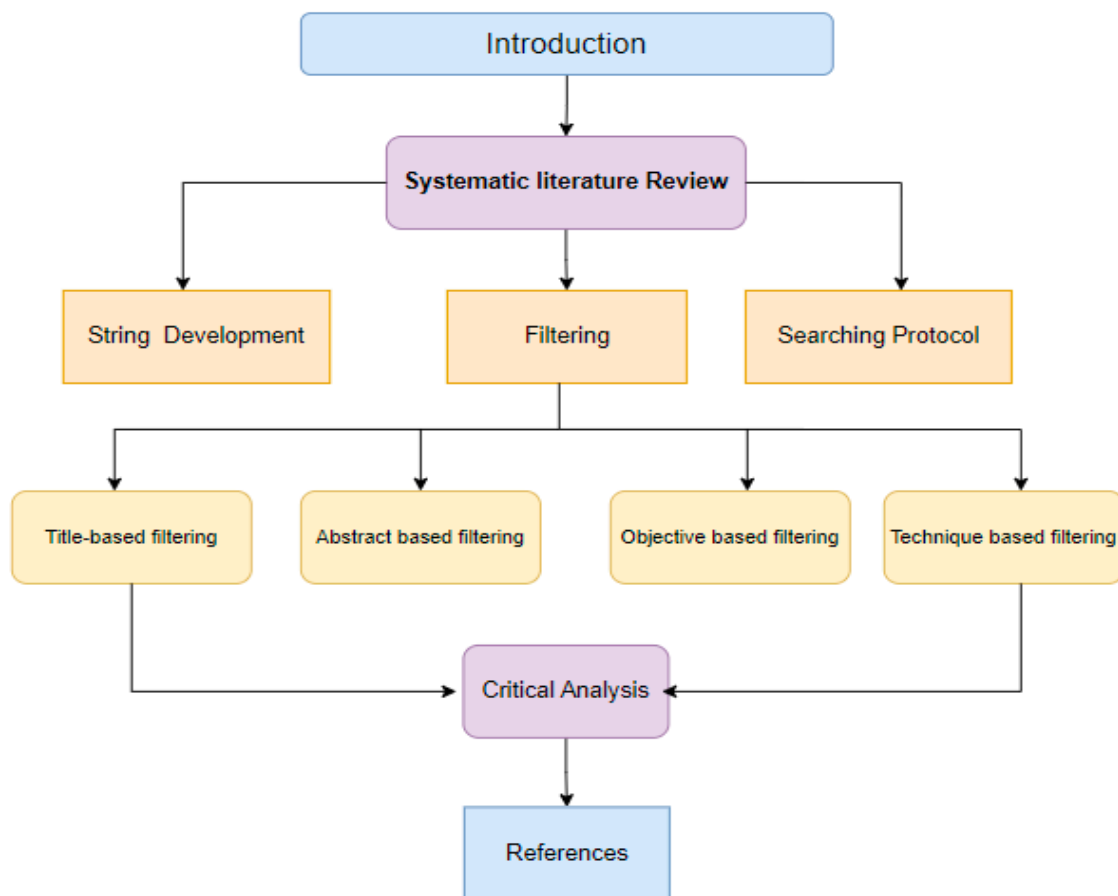

Figure 1. Organization of Paper.

With the investigation of more effective strategies and processes, accuracy rates could be raised from the currently used methods' above average levels [4]. Components of speed, In addition to reliability, homonyms and homographs in both the source and target languages must be addressed. This paper Claims that there are less resources accessible for languages that are not frequently used in informal communication, despite increased efforts being made to create resources for other formal languages [2]. Different languages are known to have their own distinctive ways of expression; for instance, in S17, Singaporeans typically talk and write in English while also utilizing some Malay and Chinese dialects, which undoubtedly contain a mix of informal languages, it's a challenge so that future research should be able to process texts in informal representations in addition to formal multilingual texts. The investigation reveals that accuracy is the most popular way to evaluate the effectiveness of sentiment analysis suggestions. However, the accuracy model only produces a single number without describing the different kinds of errors that might occur during the evaluation process, it should be noted that using this model alone might not be sufficient to guarantee that the results can be used solely as the indicator [3]. The imbalanced number of occurrences in the various classes also has a significant impact on accuracy Therefore, the effectiveness of sentiment analysis models could be accurately assessed various evaluation models, such as accuracy and precision or accuracy and f-measure, in combination.

Table 1. Summary of Survey of Multilingual Sentiment Analysis.

Years	Main Focus	Major Contribution	Enhancement in Our Paper
2021	sentiment analysis research for the Portuguese language	Describe and classify work that broadly relate to sentiment analysis. Discussing work using sentiment analysis techniques tailored to the Portuguese language and its auxiliary materials, such as natural language processing (NLP) tools, lexicons corpora, ontologies, taxonomies, and databases.[2]	The research study presents a comprehensive comparative analysis of all existing machine learning techniques and detailed critical analysis.
2021	Combinations of language for multilingual sentiment analysis.	In order to construct a multilingual sentiment analysis, this study will explore various combinations of preprocessing approaches, sentiment analysis techniques, and assessment models that have been used in the already suggested models [3].	A proper combination of models of Evaluation have been discussed for Evaluating Multilingual Sentiment Analysis.
2020	a thorough examination of sentiment analysis techniques used with non-English languages	The tools, strategies, mechanisms, and performances are covered in detail, with a focus on machine learning methods, which are being used for Multilingual Sentiment Analysis [4].	The paper have not focused on only machine translation techniques but also on other machine learning techniques used for multilingual Sentiment analysis.

Table 1 makes it evident that there are a number of gaps in the current polls. The surveys lack a systematic presentation and comprehensive critical and comparison evaluations model for multilingual sentiment analysis. The supplied surveys do not cover the newest tools for multilingual analysis, such as artificial intelligence-based systems and machine learning techniques. In order to contribute to the field and fill in the gaps in the current surveys, we have created this thorough literature review. The latest techniques and cutting-edge methodologies an evaluation models for multilingual sentiment analysis are presented in this work.

The paper is designed as following 6 sections. The Section 2 discusses the process for the systematic literature review, Comprehensive literature review is presented in Section 3, and performance analysis is presented in Section 4. Based on the critical analysis and comparison analysis. The final section, Section 6, offers conclusions after Section 5 analyses the ideal options.

Systematic Literature Review

The academic works cited in this article has been systematically reviewed. First one, a search protocol was developed and systematic searches were performed accordingly. These searchers were guided in constructing strings according to identified research questions. A search strategy was then developed to classify all searches. After search journal. In addition, the research papers Filtered by their inclusion criteria and by title, abstract, and purpose.

String Development

We have developed the strings for research work by using three synonyms for each keyword as we can see in Table 2 which shows 3 synonym for each research question word and then query for research is generated accordingly as given below.

Research question: Multilingual sentiment Analysis using deep learning

Table 2. String development.

WORDS	SYNONYM 1	SYNONYM 2	SYNONYM 3
Multilingual	Bilingual	Cross lingual	Trilingual
Sentiment	Emotion	Opinion	Sentimentalism
Analysis	Detection	Mining	—
Query NO 1	Multilingual Sentiment Analysis using deep learning		
Query NO 2	Cross lingual Sentiment Analysis using deep learning		
Query NO 3	Bilingual Sentiment Analysis using deep learning		
Query NO 4	Multilingual opinion Analysis using deep learning		
Query NO 5	Cross lingual emotion Analysis using deep learning		
Query NO 6	Multilingual Sentiment Mining using deep learning		
Query NO 7	Multilingual Opinion mining using deep learning		
Query NO 8	Cross lingual opinion mining using deep learning		
Query NO 9	Bilingual opinion mining using deep learning		

Searching Protocol

A search strategy was developed using publications published over a four-year period of time, (2020, 2021, 2022, and 2023) were chosen for search. Additionally, each keyword has three synonyms, and there are three databases (IEEE, Google Scholar, and ACM). Figure 2 presents the search strategies.

Filtering:

Figure 3 illustrates the first stage of the filtering phase, which was title-based filtering. All publications that did not address the issue at hand were removed from the databases that were chosen.
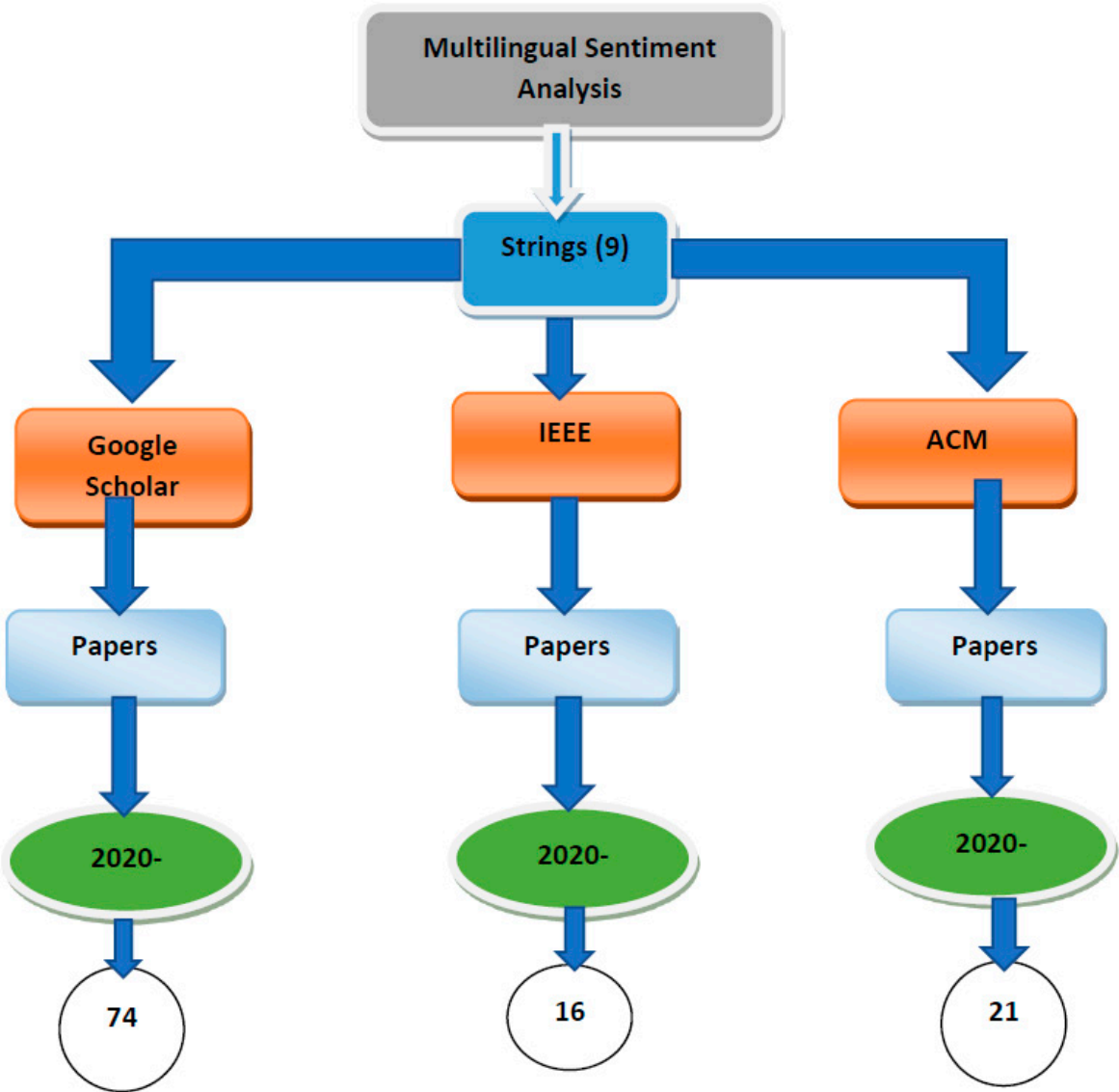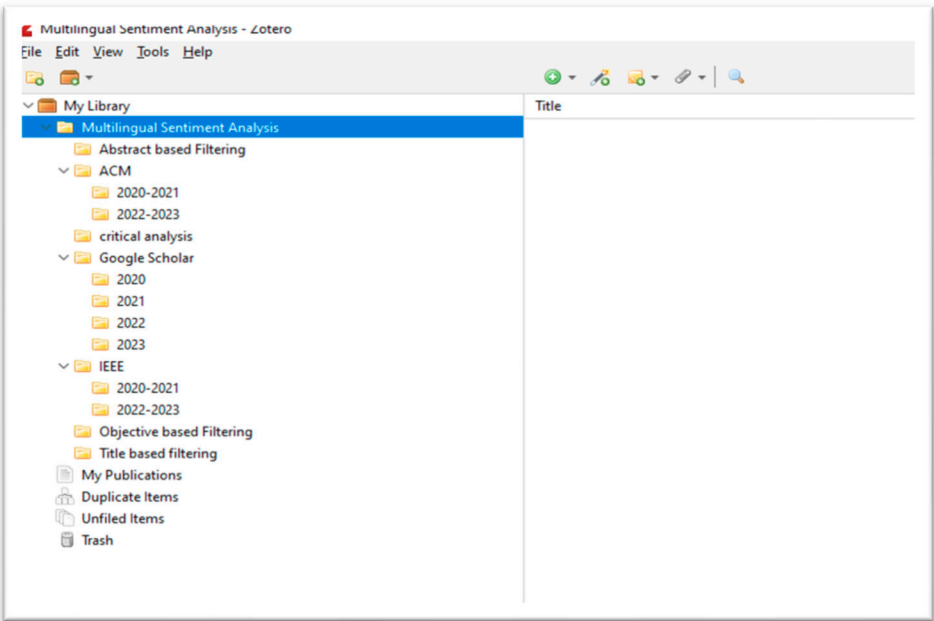
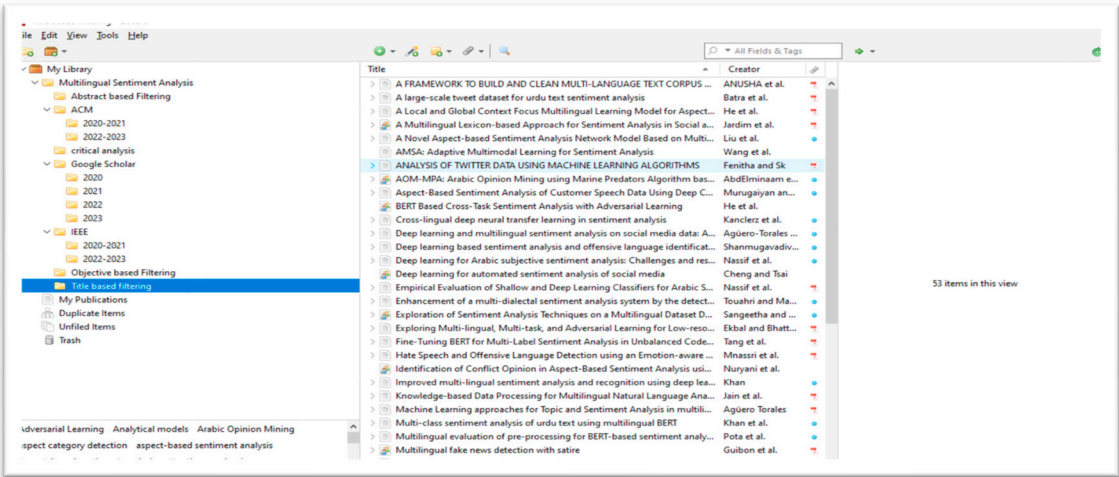**Figure 2.** Searching protocol using various Databases.

We have created folder for our research project in Zotero and then added sub collection for each database i.e.; Google Scholar, IEEE, ACM. For Every database we added sub collection folders year wise from 2020 to 2023 for storing research papers accordingly in database as shown below.

1.2.1. Title based Filtering

In first phase of filtering Papers that are relevant to title of our research question are extracted.40 papers relevant to multilingual sentiment analysis are extracted as shown in Figure 3.

We have added subfolder to our main folder in zotero as shown below. We have filtered papers stored in database on zotero on the basis of Titles related to our research question. The filtered papers on the basis of titles are added to Title based filtered Folder.

Figure 3. Selecting research articles for screening in order to determine the research's goals.

Abstract based Filtering

Abstract-based filtering was done in the second section. All papers with abstracts that were not pertinent to the issue were eliminated. And which are relevant to our research are included from all the selected databases as shown in Figure 3.

In the next phase we have filtered papers and choose those papers in which Abstract were relevant to our research and then added those relevant papers to Abstract based filtering Folder as shown below.

Table 3. Notation Table.

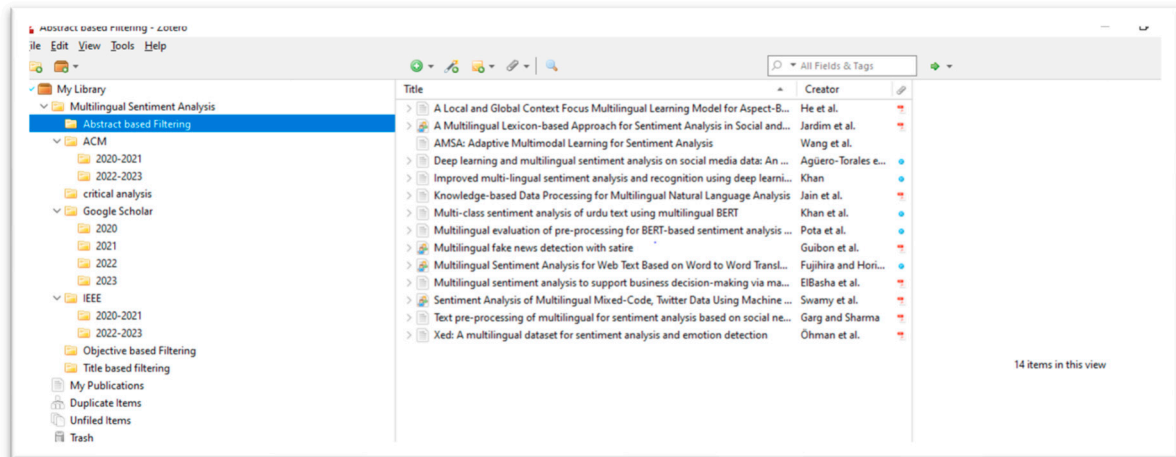| Acronyms	Definition	Acronym	Definition
DM	Decision Making	LGCF	Local global context Focus
MSA	Multilingual Sentiment Analysis	CNN	Convolutional Neural Network
E	Efficiency	BGRU	Bidirectional Gated Recurrent Unit
A	Accuracy	RF	Random Forest
FE	Feature Extraction	SVM	Support Vector Machine
ABSA	Aspect based Sentiment analysis	BERT	Bidirectional Encoder Representation from Transformer
FPE	False positive error	TVTS	train-validate-test split
MLA	Multilingual lexicon based analysis		

Objective based Filtering

As seen in the third section of the filtering, objective-based filtering was carried out. The papers were all sorted according to their objectives, as demonstrated below, and a table was created that displays the papers in accordance with those objectives. The goals of the research papers were Determined and grouped into clusters following the title and abstract-based clustering.
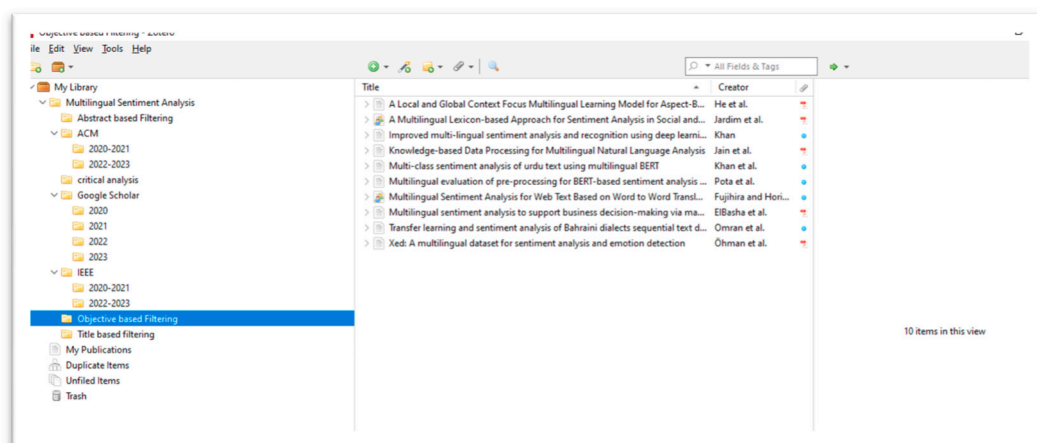

Figure 5. Objective based Filtering.

The main aim of all selected research papers is to perform multilingual sentiment analysis efficiently. Table 4 shows this objective-based screening, by defining objectives of each paper by analyzing them. The categories of objectives are as follow and their definition is in Table 3.

Table 4. Objective-based screening.

Ref.	Performance	Classification	FE	DM	ABSA	E	A	MSA
[5]					✓	✓		✓
[6]		✓	✓			✓		✓
[7]						✓	✓	✓
[1]			✓		✓		✓	✓
[8]	✓						✓	✓
[9]	✓	✓	✓					✓
[10]			✓	✓		✓		✓
[11]				✓			✓	✓
[12]	✓				✓			
[13]	✓						✓	

Techniques

The main aim of all the papers selected is to design a model for multilingual sentiment analysis using machine learning techniques.

Table 5. Summary of Methodologies used for multilingual sentiment analysis.

Ref	Scheme	Methodology
[5]	LGCF,CNN,BGRU	This paper design a new model called LGCF for aspect- predicated Sentiment analysis, by combining CNN and BGRU for the Global terrain Focus. Global terrain Focus mainly consists of three corridor BGRU, CNN, and Layer Normalization (LN).
[6]	lexicon-based sentiment analysis algorithm	This paper described an unsupervised automatic system of multilingual dictionary- predicated sentiment analysis algorithm, where we used a dictionary- predicated approach. This method involves employing a dictionary that has been pre-defined and contains both positive and negative words.
[7]	mBert,word embedding,gpt	Four text representations word were used i.e. n- grams, char n-- trained fast Text and BERT word embeddings to train our classifiers. These models were trained using different datasets to evaluate the results. Adapting shows that the suggested mBERT model with BERT pre-trained word embeddings outperforms deep knowledge. Machine knowledge and rule- predicated classifiers.

[1]	XGBoost classifier,RF	In the proposed work, emotion recognition across languages is developed using Urdu, Italian, English, and German. The most common audio point is used to pull the features. Known as Mel frequency Cepstral Portions. XGBoost classifier and Rf is used for analysis.
[8]	HEGLE, Ker_Rad_BF,CNN	Multilingual data processing is proposed in this paper using point birth with type using deep knowledge architectures. The reused data has been pulled using Histogram Equalization predicated Global Original Entropy(HEGLE) and classified using Kernel- predicated Radial base Function(Ker_Rad_BF).
[9]	BERT,NN,Word embedding	The novel Arabic sentiment analysis model presented in this study is trained on the English sentiment analysis model. The goal is to create a framework that makes use of languages with large data sets to build a comprehensive model. A word embedding fashion in addition to machine knowledge model approach are both executed to translate, count point birth and resource conditions for sentiment analysis.
[10]	NN,Word embedding	In this disquisition, We have developed a multilingual sentiment analysis system that relies on word-to-word paraphrasing using a sentiment dictionary in any given native language. This system consists of three main phases: morphological analysis of the text, determining the sentiment of each word using a sentiment dictionary, and aggregating the sentiments of individual words to determine the overall sentiment of the text. To evaluate the effectiveness of our system, we conducted a sentiment analysis trial on tweets written in English, German, French, and Spanish.
[11]	BERT,SVM	In this paper Bert and SVM machine knowledge ways were used for Multilingual Sentiment analysis.
[12]	BERT	In this paper number of trials, performed by a state- of- the- art type model (BERT), are designed, to estimate multitudinous presently available pre-processing techniques for tweets and the statistical impact of those techniques on sentiment analysis results.
[10]	k-fold cross-validation,TVTS	In this paper to use and convey the information learned from evaluating the product, a pre-trained Long Short Term Memory model was developed. Reviews in Bahraini cants to perform sentiment analysis on a small dataset of movie commentary in the same cants.

Detailed Literature

This paper proposed a model for multilingual literacy grounded on original and global environment concentrated interactive literacy, videlicet LGCF. By Comparing it with being models, this model can successfully learn the correlation between the goal aspects and the original environment, as well as the correlation between global environment and target aspects contemporaneously. Moreover, this technique can effectively dissect both Chinese and English review. Trials conducted on three Chinese standard datasets which show that LGCF improves

performance as compared to several being situations. And a satisfying performance has been achieved [5].

This paper presents an unsupervised robotic sentiment *analysis* algorithm grounded on multilingual vocabulary, with a wordbook- grounded approach, is described. The devised schemes were implemented and tested in the sentiment analysis of drug-related publications on a digital tourism platform. The obtained results demonstrate the effectiveness of the approach, indicating a high level of accuracy in categorizing publications written in four different languages [6].

This paper introduced a new multi-class Urdu dataset grounded on stoner commentary for sentiment analysis. This dataset has been collected from colorful fields similar as food and potables, pictures and plays, software and operations, politics and sports [7]. We OK - tuned Multilingual BERT (mBERT) for Urdu sentiment analysis and used four textbook representations word n- gram, housekeeper n- gram, pre-trained presto Text and BERT word embedding help our classifiers learn. For evaluation purposes, we used two separate datasets to train these models.

Urdu, Italian, English, German, and English are used to establish cross-language emotion recognition in the proposed work. The most often utilized audio point, known as MFCC (Mel frequency Cepstral Portions), uproots the features. The experimental findings demonstrated that the suggested deep literacy model on the URDU dataset with a delicacy of 91.25 employing Random Forest (RF) and XGBoost bracket offers promising outcomes [8].

We express present results to the multilingual sentiment analysis issue in this exploration composition by enforcing algorithms, and we compare perfection factors to discover the optimum option for multilingual sentiment analysis [9].

A word embedding fashion is enforced [1] in addition to the machine literacy model approach, both for restatement, barring point birth and coffers needed for sentiment analysis. The developed model is trained in such a way that associations can use it to understand client stations about their products.

We conduct a sentiment bracket trial] during the trial, we evaluated the performance of our classifier on tweets written in English, German, French, and Spanish. To assess its effectiveness, we compared our classifier with previous classifiers using evaluation metrics such as "accuracy," "precision," "recall," and "F1 score." The experimental results indicate that our classifier is suitable for multilingual sentiment analysis, as its performance remains consistent regardless of language differences [11].

XED is a new and advanced dataset that provides a new challenge in accurate emotion recognition with preliminarily unapproachable language content. Maybe the biggest donation of all is that, for the first time, numerous resource-poor languages sentiment datasets can be used in other possible downstream operations [12].

Advanced Bracket Model (BERT), are designed to estimate numerous of the presently available operations for tweet preprocessing, in terms of the statistical significance of their impact on sentiment analysis performance. In addition, data available in two languages, videlicet English and Italian, are considered to assess language dependence [13].

The end of this exploration is to give an applicable multilingual deep literacy short-term memory model for assaying resembling datasets of English cants, ultramodern Standard Arabic and Bahraini cants that differ in verbal features. A short-term memory model with pre-training was created, to utilize and apply the understanding obtained from analyzing product reviews in Bahraini cants to sentiment analysis on a small dataset of movie reviews in the same cants. [14]. In our survey on multilingual sentiment analysis using deep learning, our research builds upon foundational insights presented in [21–36]

Performance Analysis

LGCF has limited ability to capture context in long sentences: LGCF model processes text sequentially, which may lead to a loss of context in long sentences. This can impact the model's ability to accurately identify and classify sentiment for different aspects in such sentences [5].

In Dependence on good lexicons: The MLA approach's effectiveness is heavily reliant on the caliber of the lexicons used for mood analysis. The accuracy of the approach may be impacted by the quality of the lexicons, which can vary greatly based on the language and the domain. Contextual understanding is limited because of the MLA approach's presumption that words have a set polarity regardless of their context. When the same word has a different sentiment based on the context, this can result in inaccurate sentiment analysis results [6].

In this paper limited ability to capture long-term dependencies: Despite being able to capture context in both ways thanks to its bidirectional architecture, BERT may still have trouble capturing long-term dependencies in text. This may affect how well it performs jobs that call for comprehending intricate relationships between words and phrases [7].

Machine translation of languages with limited resources into languages with abundant resources has also been studied. The development of distinct monolingual Sentiment analysis models for many languages is no longer necessary because to improvements in statistical or neural machine translation systems. The multilingual data set is translated into English using neural machine translation, which is then labelled as the English language model. If the source and target languages are the same, word-to-word translation can be employed to reduce computing costs. However, the value of machine translation cannot be overstated because the accuracy of texts written in non-normative languages is not always dependable and can even introduce more noise. [8].

In SVM can be computationally precious and slow when working with large- scale datasets, which can limit its practicality for some operations. Limited performance on imbalanced datasets SVM may struggle to perform well on imbalanced datasets, where the number of exemplifications in one class is much lesser than the other BERT's bidirectional armature allows it to capture environment from both directions, it may still struggle with landing long- term dependences in textbook. This can impact its performance on tasks that bear understanding complex connections between words and expressions [9]. Moreover, [37–50] serve as pillars of knowledge, contributing to the foundation upon which our exploration and analysis are grounded

Critical Analysis

BERT's performance can vary significantly depending on the quantum and quality of training data available for a particular language. This can limit its effectiveness in low- resource languages or disciplines [10].

Table 6. Analysis.

Reference	Effort Year	Methodology	Shortcoming
[5]	2022	LGCF	LGCF model processes text sequentially, which may lead to a loss of context in long sentences. This can impact the model's ability to accurately identify and classify sentiment for different aspects in such sentences [5].
[6]	2021	Lexicon-based Algorithm	The accuracy of the approach may be impacted by the quality of the lexicons, which can vary greatly based on the language and the domain. Contextual understanding is limited because of the MLA approach's presumption that words have a set polarity regardless of their context. When the same word has a different sentiment based on the context, this can result in inaccurate sentiment analysis results[14].

[7]	2022	mBert,word Embedding	Despite being able to capture context in both ways thanks to its bidirectional architecture, BERT may still have trouble capturing long-term dependencies in text. This may affect how well it performs jobs that call for comprehending intricate relationships between words and phrases [15].
[8]	2020	NN,Word embedding	If the source and target languages are the same, word-to-word translation can be used to save on computational expenses. The multilingual data set is translated into English using neural machine translation, which is then labelled as the English language model. [16]
[9]	2020	SVM,Bert	SVM can be computationally precious and slow when working with large- scale datasets, which can limit its practicality for some operations, BERT's bidirectional armature allows it to capture environment from both directions, it may still struggle with landing long-term dependences in textbook. This can impact its performance on tasks that bear understanding complex connections between words and expressions [17] [18].
[10]	2021	BERT	BERT's performance can vary significantly depending on the quantum and quality of training data available for a particular language. This can limit its effectiveness in low- resource languages or disciplines.[19]

Research Gaps

The following Table shows the gaps in previous studies and their solution are also discussed below.

Table 7. Research gaps.

Ref.	Research Gaps	Solution
[6]	Inability to recognize nuances	Developing more sophisticated lexicons as well as using machine learning techniques that can increase the precision of sentiment analysis in data from social and cultural information systems[20].
[5]	The availability of annotated data is limited.	Exploring methods for utilizing Unannotated data or the artificial data production can help ABSA models to perform in a better way for languages with limited resources. [14].

[7]	Only Urdu literature is used to evaluate the suggested technique.	A more thorough evaluation of the model's effectiveness may result from further testing it on additional languages [20].
[11]	Limited exploration of multilingual models	Sentiment analysis and emotion identification can be done more quickly and accurately by implementing multilingual models that can handle numerous languages at once [1].
[8] [9]	Lack of analysis of model performance on specific business domains	By providing transparency in the model selection process and examining model performance on specific business domains, the suggested methodology for sentiment analysis to support corporate decision-making using machine learning models becomes more valuable [1],[16].

Dataset

Table 8. Dataset description.

References	Dataset Description
[6]	The data set for this work is constructed from the comments collected for each existing Cuscaria, it focuses on the data kept in the Cuscarias cultural and social information system.
[7]	They gathered information from a variety of genres, including films, dramas from Pakistan and India, TV debates, cooking programs, politicians and Pakistani political parties, sports, software, blogs and forums, and gadgets.
[8]	Three categories of dataset are used, the first is actual data from numerous actual experiments. The second kind is synthetic data, which is made in an effort to mimic actual trends. The third category, which is used for presentation and visualization, is for toy datasets.
[9]	In this study, review and tweet corpora that are both open to the public were used. The first corpus is made up of reviews in English, and the second one is made up of reviews in Arabic. The SemEval-2016 Challenge Task 5 calls for the use of these data sets.
[11]	The parallel movie subtitle corpus OPUS (Lison and Tiedemann, 2016) was built using opensubtitles.org as a multi-domain proxy.
[12]	The dataset made available at SemEval 2017 is used for English language, specifically the one linked with (Rosenthal et al., 2017) to task 4A. Additionally, the SENTIPOLC 2016 (SENTIment Polarity Classification) dataset for Italian is examined. It was presented at the EVALITA 2016 conference, which evaluated NLP and voice tools for Italian.
[13].	A corpus of Sudanese Dialect Arabic (SDA). SDA is a political-focused lexical resource comprising of 5456 tweets that were gathered using the Twitter API. Another corpus of 40,000 tweets. Egyptian dialects and MSA are mixed together in the 40,000 tweets. These tweets, which cover subjects like proverbs, poetry, caustic jokes, social issues, politics, health, sports, and product opinions, were gathered using the Twitter API.

Conclusion

This survey paper provided a comprehensive overview of the approaches, challenges, and future directions in multilingual sentiment analysis. We discussed the challenges and limitations of multilingual sentiment analysis. Data scarcity in low-resource languages, domain adaptation, cross-lingual transfer, and language-specific nuances pose significant challenges in achieving accurate sentiment analysis across languages. This paper highlighted the need for carefully curated multilingual datasets, language-specific preprocessing techniques, and fine-tuning strategies to address these challenges.

The field continues to evolve as researchers strive to improve sentiment analysis capabilities across languages. By understanding the existing techniques and addressing the challenges, we can achieve more accurate and robust multilingual sentiment analysis, enabling effective decision-making and understanding of sentiment in diverse linguistic contexts.

References

1. A. Khan, "Improved multi-lingual sentiment analysis and recognition using deep learning," J. Inf. Sci., p. 01655515221137270, 2023.
2. D. A. Pereira, "A survey of sentiment analysis in the Portuguese language," Artif. Intell. Rev., vol. 54, no. 2, pp. 1087–1115, 2021.
3. N. A. S. Abdullah and N. I. A. Rusli, "Multilingual Sentiment Analysis: A Systematic Literature Review." Pertanika J. Sci. Technol., vol. 29, no. 1, 2021.
4. S. Sagnika, A. Pattanaik, B. S. P. Mishra, and S. K. Meher, "A review on multi-lingual sentiment analysis by machine learning methods," J. Eng. Sci. Technol. Rev., vol. 13, no. 2, pp. 154–166, Jan. 2020, doi: 10.25103/jestr.132.19.
5. J. He, A. Wumaier, Z. Kadeer, W. Sun, X. Xin, and L. Zheng, "A Local and Global Context Focus Multilingual Learning Model for Aspect-Based Sentiment Analysis," IEEE Access, vol. 10, pp. 84135–84146, 2022, doi: 10.1109/ACCESS.2022.3197218.
6. S. Jardim, C. Mora, and T. Santana, "A Multilingual Lexicon-based Approach for Sentiment Analysis in Social and Cultural Information System Data," in 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), Jun. 2021, pp. 1–6. doi: 10.23919/CISTI52073.2021.9476631.
7. L. Khan, A. Amjad, N. Ashraf, and H.-T. Chang, "Multi-class sentiment analysis of urdu text using multilingual BERT," Sci. Rep., vol. 12, no. 1, p. 5436, 2022.
8. D. K. Jain, Y. G.-M. Eyre, A. Kumar, B. B. Gupta, and K. Kotecha, "Knowledge-based Data Processing for Multilingual Natural Language Analysis," ACM Trans. Asian Low-Resour. Lang. Inf. Process..
9. S. ElBasha, A. Elhawil, and N. Drawil, "Multilingual sentiment analysis to support business decision-making via machine learning models," Univ Tripoli Tripoli Libya Tech Rep, 2021.
10. K. Fujihira and N. Horibe, "Multilingual Sentiment Analysis for Web Text Based on Word to Word Translation," in 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI), Sep. 2020, pp. 74–79. doi: 10.1109/IIAI-AAI50415.2020.00025.
11. E. Öhman, M. Pàmies, K. Kajava, and J. Tiedemann, "Xed: A multilingual dataset for sentiment analysis and emotion detection," ArXiv Prepr. ArXiv201101612, 2020.
12. M. Pota, M. Ventura, H. Fujita, and M. Esposito, "Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets," Expert Syst. Appl., vol. 181, p. 115119, 2021.
13. T. M. Omran, B. T. Sharef, C. Grosan, and Y. Li, "Transfer learning and sentiment analysis of Bahraini dialects sequential text data using multilingual deep learning approach," Data Knowl. Eng., vol. 143, p. 102106, 2023.
14. L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," IEEE Access, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.
15. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," ArXiv Prepr. ArXiv181004805, 2018.
16. H. Rahab, H. Haouassi, M. E. H. Souidi, A. Bakhouche, R. Mahdaoui, and M. Bekhouche, "A Modified Binary Rat Swarm Optimization Algorithm for Feature Selection in Arabic Sentiment Analysis," Arab. J. Sci. Eng., Dec. 2022, doi: 10.1007/s13369-022-07466-1.
17. B. Li, Y. Weng, Q. Song, B. Sun, and S. Li, "Continuing Pre-trained Model with Multiple Training Strategies for Emotional Classification," in Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis, 2022, pp. 233–238.
18. S. Deb and A. K. Chanda, "Comparative analysis of contextual and context-free embeddings in disaster prediction from Twitter data," Mach. Learn. Appl., vol. 7, p. 100253, 2022.

19. R. Ortega-Bueno, P. Rosso, and J. E. M. Pagola, "Multi-view informed attention-based model for Irony and Satire detection in Spanish variants," Knowl.-Based Syst., vol. 235, p. 107597, 2022.

20. J. Y.-L. Chan, K. T. Bea, S. M. H. Leow, S. W. Phoong, and W. K. Cheng, "State of the art: a review of sentiment analysis based on sequential transfer learning," Artif. Intell. Rev., vol. 56, no. 1, pp. 749–780, 2023.

21. Kok, S. H., Azween, A., & Jhanjhi, N. Z. (2020). Evaluation metric for crypto-ransomware detection using machine learning. Journal of Information Security and Applications, 55, 102646.

22. Shafiq, M., Ashraf, H., Ullah, A., Masud, M., Azeem, M., Jhanjhi, N. Z., & Humayun, M. (2021). Robust Cluster-Based Routing Protocol for IoT-Assisted Smart Devices in WSN. Computers, Materials & Continua, 67(3).

23. Lim, M., Abdullah, A., Jhanjhi, N. Z., & Supramaniam, M. (2019). Hidden link prediction in criminal networks using the deep reinforcement learning technique. Computers, 8(1), 8.

24. Sennan, S., Somula, R., Luhach, A. K., Deverajan, G. G., Alnumay, W., Jhanjhi, N. Z., ... & Sharma, P. (2021). Energy efficient optimal parent selection based routing protocol for Internet of Things using firefly optimization algorithm. Transactions on Emerging Telecommunications Technologies, 32(8), e4171.

25. Hussain, K., Hussain, S. J., Jhanjhi, N. Z., & Humayun, M. (2019, April). SYN flood attack detection based on bayes estimator (SFADBE) for MANET. In 2019 International Conference on Computer and Information Sciences (ICCIS) (pp. 1-4). IEEE.

26. Lim, M., Abdullah, A., & Jhanjhi, N. Z. (2021). Performance optimization of criminal network hidden link prediction model with deep reinforcement learning. Journal of King Saud University-Computer and Information Sciences, 33(10), 1202-1210.

27. Gaur, L., Singh, G., Solanki, A., Jhanjhi, N. Z., Bhatia, U., Sharma, S., ... & Kim, W. (2021). Disposition of youth in predicting sustainable development goals using the neuro-fuzzy and random forest algorithms. Human-Centric Computing and Information Sciences, 11, NA.

28. Kumar, T., Pandey, B., Mussavi, S. H. A., & Zaman, N. (2015). CTHS based energy efficient thermal aware image ALU design on FPGA. Wireless Personal Communications, 85, 671-696.

29. Nanglia, S., Ahmad, M., Khan, F. A., & Jhanjhi, N. Z. (2022). An enhanced Predictive heterogeneous ensemble model for breast cancer prediction. Biomedical Signal Processing and Control, 72, 103279.

30. Gaur, L., Afaq, A., Solanki, A., Singh, G., Sharma, S., Jhanjhi, N. Z., ... & Le, D. N. (2021). Capitalizing on big data and revolutionary 5G technology: Extracting and visualizing ratings and reviews of global chain hotels. Computers and Electrical Engineering, 95, 107374.

31. Diwaker, C., Tomar, P., Solanki, A., Nayyar, A., Jhanjhi, N. Z., Abdullah, A., & Supramaniam, M. (2019). A new model for predicting component-based software reliability using soft computing. IEEE Access, 7, 147191-147203.

32. Shahid, H., Ashraf, H., Javed, H., Humayun, M., Jhanjhi, N. Z., & AlZain, M. A. (2021). Energy optimised security against wormhole attack in iot-based wireless sensor networks. Comput. Mater. Contin, 68(2), 1967-81.

33. Wassan, S., Chen, X., Shen, T., Waqar, M., & Jhanjhi, N. Z. (2021). Amazon product sentiment analysis using machine learning techniques. Revista Argentina de Clínica Psicológica, 30(1), 695.

34. Almusaylim, Z. A., Zaman, N., & Jung, L. T. (2018, August). Proposing a data privacy aware protocol for roadside accident video reporting service using 5G in Vehicular Cloud Networks Environment. In 2018 4th International conference on computer and information sciences (ICCOINS) (pp. 1-5). IEEE.

35. Ghosh, G., Verma, S., Jhanjhi, N. Z., & Talib, M. N. (2020, December). Secure surveillance system using chaotic image encryption technique. In IOP conference series: materials science and engineering (Vol. 993, No. 1, p. 012062). IOP Publishing.

36. Humayun, M., Alsaqer, M. S., & Jhanjhi, N. (2022). Energy optimization for smart cities using iot. Applied Artificial Intelligence, 36(1), 2037255.

37. E. Ndashimye, N. I. Sarkar, and S. K. Ray, "A Multi-criteria based handover algorithm for vehicle-to-infrastructure communications," Computer Networks, vol. 185, no. 202152, Article ID 107652, 2020

38. Ray, S. K., Pawlikowski, K., & Sirisena, H. (2009). A fast MAC-layer handover for an IEEE 802.16 e-based WMAN. In AccessNets: Third International Conference on Access Networks, AccessNets 2008, Las Vegas, NV, USA, October 15-17, 2008. Revised Papers 3 (pp. 102-117). Springer Berlin Heidelberg.

39. Srivastava, R. K., Ray, S., Sanyal, S., & Sengupta, P. (2011). Mineralogical control on rheological inversion of a suite of deformed mafic dykes from parts of the Chottanagpur Granite Gneiss Complex of eastern India. Dyke Swarms: Keys for Geodynamic Interpretation: Keys for Geodynamic Interpretation, 263-276.

40. Ray, S. K., Sinha, R., & Ray, S. K. (2015, June). A smartphone-based post-disaster management mechanism using WiFi tethering. In 2015 IEEE 10th conference on industrial electronics and applications (ICIEA) (pp. 966-971). IEEE.

41. Chaudhuri A, Ray S (2015) Antiproliferative activity of phytochemicals present in aerial parts aqueous extract of Ampelocissus latifolia (Roxb.) planch. on apical meristem cells. Int J Pharm Bio Sci 6(2):99–108

42. Hossain, A., Ray, S. K., & Sinha, R. (2016, December). A smartphone-assisted post-disaster victim localization method. In 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 1173-1179). IEEE.

43. Airehrour, D., Gutierrez, J., & Ray, S. K. (2018). A trust-based defence scheme for mitigating blackhole and selective forwarding attacks in the RPL routing protocol. Journal of Telecommunications and the Digital Economy, 6(1), 41-49.

44. Ray, S. K., Ray, S. K., Pawlikowski, K., McInnes, A., & Sirisena, H. (2010, April). Self-tracking mobile station controls its fast handover in mobile WiMAX. In 2010 IEEE Wireless Communication and Networking Conference (pp. 1-6). IEEE.

45. Dey, K., Ray, S., Bhattacharyya, P. K., Gangopadhyay, A., Bhasin, K. K., & Verma, R. D. (1985). Salicyladehyde 4-methoxybenzoylhydrazone and diacetylbis (4-methoxybenzoylhydrazone) as ligands for tin, lead and zirconium. J. Indian Chem. Soc.;(India), 62(11).

46. Airehrour, D., Gutierrez, J., & Ray, S. K. (2017, November). A testbed implementation of a trust-aware RPL routing protocol. In 2017 27th International Telecommunication Networks and Applications Conference (ITNAC) (pp. 1-6). IEEE.

47. Ndashimye, E., Sarkar, N. I., & Ray, S. K. (2016, August). A novel network selection mechanism for vehicle-to-infrastructure communication. In 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech) (pp. 483-488). IEEE.

48. Ndashimye, E., Sarkar, N. I., & Ray, S. K. (2020). A network selection method for handover in vehicle-to-infrastructure communications in multi-tier networks. Wireless Networks, 26, 387-401.

49. Siddiqui, F. J., Ashraf, H., & Ullah, A. (2020). Dual server based security system for multimedia Services in Next Generation Networks. Multimedia Tools and Applications, 79, 7299-7318.

50. Shahid,H.,Ashraf,H.,Ullah,A.,Band,S.S.&Elnaffar,S.Wormholeattackmitigationstrategiesandtheirimpact onwirelesssensornetworkperformance: Aliteraturesurvey. ternationalJournalofCommunicationSystems35, e5311(2022). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.5311. https://onlinelibrary.wiley.com/doi/pdf/10.1002/dac.5311

51. Zamir, U. B., Masood, H., Jamil, N., Bahadur, A., Munir, M., Tareen, P., ... & Ashraf, H. (2015, July). The relationship between sea surface temperature and chlorophyll-a concentration in Arabian Sea. In Biological Forum–An International Journal (Vol. 7, No. 2, pp. 825-834).